



**HAL**  
open science

# Comparaison d'une analyse automatique d'IRM cérébrale par PIXYL versus imagerie de fusion dans le suivi de la sclérose en plaques

Jean-Loup Méreaux

► **To cite this version:**

Jean-Loup Méreaux. Comparaison d'une analyse automatique d'IRM cérébrale par PIXYL versus imagerie de fusion dans le suivi de la sclérose en plaques. Médecine humaine et pathologie. 2021. dumas-03516897

**HAL Id: dumas-03516897**

**<https://dumas.ccsd.cnrs.fr/dumas-03516897v1>**

Submitted on 7 Jan 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

**UFR DE SANTÉ DE ROUEN NORMANDIE**

**ANNÉE 2021**

**N°**

**THÈSE POUR LE  
DOCTORAT EN MÉDECINE**

(Diplôme d'État)

Par

MÉREAUX Jean-Loup

NÉ LE 27 AVRIL 1992 À PARIS

PRÉSENTÉE ET SOUTENUE PUBLIQUEMENT LE 18 OCTOBRE 2021

**COMPARAISON D'UNE ANALYSE AUTOMATIQUE D'IRM  
CÉRÉBRALE PAR PIXYL VERSUS IMAGERIE DE FUSION DANS LE  
SUIVI DE LA SCLÉROSE EN PLAQUES**

PRESIDENT DE JURY :      Professeur David MALTÊTE

DIRECTEUR DE THESE :    Docteur Bertrand BOURRE

MEMBRES DU JURY :      Professeur Emmanuel GERARDIN

Docteur Lou GRANGEON



ANNEE UNIVERSITAIRE 2020 - 2021

U.F.R. SANTÉ DE ROUEN

-----

DOYEN : **Professeur Benoît VEBER**

ASSESEURS : **Professeur Loïc FAVENNEC**  
**Professeur Agnès LIARD**  
**Professeur Guillaume SAVOYE**

I - MEDECINE

**PROFESSEURS DES UNIVERSITES – PRATICIENS HOSPITALIERS**

Mr Frédéric ANSELME	HCN	Cardiologie
Mme Gisèle APTER	Havre	Pédopsychiatrie
Mme Isabelle AUQUIT AUCKBUR	HCN	Chirurgie plastique
Mr Jean-Marc BASTE	HCN	Chirurgie Thoracique
Mr Fabrice BAUER	HCN	Cardiologie
Mme Soumeya BEKRI	HCN	Biochimie et biologie moléculaire
Mr Ygal BENHAMOU	HCN	Médecine interne
Mr Jacques BENICHOU	HCN	Bio statistiques et informatique médicale
Mr Olivier BOYER	UFR	Immunologie
Mme Sophie CANDON	HCN	Immunologie
Mr François CARON	HCN	Maladies infectieuses et tropicales
Mr Philippe CHASSAGNE	HCN	Médecine interne (gériatrie)
Mr Moïse COEFFIER	HCN	Nutrition
Mr Vincent COMPERE	HCN	Anesthésiologie et réanimation chirurgicale
Mr Jean-Nicolas CORNU	HCN	Urologie
Mr Antoine CUVELIER	HB	Pneumologie
Mr Jean-Nicolas DACHER	HCN	Radiologie et imagerie médicale
Mr Stéfan DARMONI	HCN	Informatique médicale et techniques de communication
Mr Pierre DECHELOTTE	HCN	Nutrition
Mr Stéphane DERREY	HCN	Neurochirurgie

Mr Frédéric <b>DI FIORE</b>	CHB	Cancérologie
Mr Fabien <b>DOGUET</b>	HCN	Chirurgie Cardio Vasculaire
Mr Jean <b>DOUCET</b>	SJ	Thérapeutique - Médecine interne et gériatrie
Mr Bernard <b>DUBRAY</b>	CHB	Radiothérapie
Mr Frank <b>DUJARDIN</b>	HCN	Chirurgie orthopédique - Traumatologique
Mr Fabrice <b>DUPARC</b>	HCN	Anatomie - Chirurgie orthopédique et traumatologique
Mr Eric <b>DURAND</b>	HCN	Cardiologie
Mr Bertrand <b>DUREUIL</b>	HCN	Anesthésiologie et réanimation chirurgicale
Mme Hélène <b>ELTCHANINOFF</b>	HCN	Cardiologie
Mr Manuel <b>ETIENNE</b>	HCN	Maladies infectieuses et tropicales
Mr Thierry <b>FREBOURG</b>	UFR	Génétique
Mr Pierre <b>FREGER</b> ( <i>surnombre</i> )	HCN	Anatomie - Neurochirurgie
Mr Jean François <b>GEHANNO</b>	HCN	Médecine et santé au travail
Mr Emmanuel <b>GERARDIN</b>	HCN	Imagerie médicale
Mme Priscille <b>GERARDIN</b>	HCN	Pédopsychiatrie
M. Guillaume <b>GOURCEROL</b>	HCN	Physiologie
Mr Dominique <b>GUERROT</b>	HCN	Néphrologie
Mme Julie <b>GUEUDRY</b>	HCN	Ophthalmologie
Mr Olivier <b>GUILLIN</b>	HCN	Psychiatrie Adultes
Mr Claude <b>HOUDAYER</b>	HCN	Génétique
Mr Fabrice <b>JARDIN</b>	CHB	Hématologie
Mr Luc-Marie <b>JOLY</b>	HCN	Médecine d'urgence
Mr Pascal <b>JOLY</b>	HCN	Dermato – Vénérologie
Mme Bouchra <b>LAMIA</b>	Havre	Pneumologie
Mme Annie <b>LAQUERRIERE</b>	HCN	Anatomie et cytologie pathologiques
Mr Vincent <b>LAUDENBACH</b>	HCN	Anesthésie et réanimation chirurgicale
Mr Hervé <b>LEFEBVRE</b>	HB	Endocrinologie et maladies métaboliques
Mr Thierry <b>LEQUERRE</b>	HCN	Rhumatologie
Mme Anne-Marie <b>LEROI</b>	HCN	Physiologie
Mr Hervé <b>LEVESQUE</b>	HCN	Médecine interne
Mme Agnès <b>LIARD-ZMUDA</b>	HCN	Chirurgie Infantile
Mr Pierre Yves <b>LITZLER</b>	HCN	Chirurgie cardiaque
M. David <b>MALTETE</b>	HCN	Neurologie
Mr Christophe <b>MARGUET</b>	HCN	Pédiatrie
Mme Isabelle <b>MARIE</b>	HCN	Médecine interne
Mr Jean-Paul <b>MARIE</b>	HCN	Oto-rhino-laryngologie
Mr Loïc <b>MARPEAU</b>	HCN	Gynécologie - Obstétrique

Mr Stéphane <b>MARRET</b>	HCN	Pédiatrie
Mme Véronique <b>MERLE</b>	HCN	Epidémiologie
Mr Pierre <b>MICHEL</b>	HCN	Hépto-gastro-entérologie
M. Benoit <b>MISSET</b> ( <i>détachement</i> )	HCN	Réanimation Médicale
Mr Marc <b>MURAINÉ</b>	HCN	Ophthalmologie
Mr Christian <b>PFISTER</b>	HCN	Urologie
Mr Jean-Christophe <b>PLANTIER</b>	HCN	Bactériologie - Virologie
Mr Didier <b>PLISSONNIER</b>	HCN	Chirurgie vasculaire
Mr Gaëtan <b>PREVOST</b>	HCN	Endocrinologie
Mr Jean-Christophe <b>RICHARD</b> ( <i>détachement</i> )	HCN	Réanimation médicale - Médecine d'urgence
Mr Vincent <b>RICHARD</b>	UFR	Pharmacologie
Mme Nathalie <b>RIVES</b>	HCN	Biologie du développement et de la reproduction
Mr Horace <b>ROMAN</b> ( <i>détachement</i> )	HCN	Gynécologie - Obstétrique
Mr Jean-Christophe <b>SABOURIN</b>	HCN	Anatomie – Pathologie
Mr Mathieu <b>SALAUN</b>	HCN	Pneumologie
Mr Guillaume <b>SAVOYE</b>	HCN	Hépto-gastrologie
Mme Céline <b>SAVOYE-COLLET</b>	HCN	Imagerie médicale
Mme Pascale <b>SCHNEIDER</b>	HCN	Pédiatrie
Mr Lilian <b>SCHWARZ</b>	HCN	Chirurgie Viscérale et Digestive
Mr Michel <b>SCOTTE</b>	HCN	Chirurgie digestive
Mme Fabienne <b>TAMION</b>	HCN	Thérapeutique
Mr Luc <b>THIBERVILLE</b>	HCN	Pneumologie
Mr Hervé <b>TILLY</b> ( <i>surnombre</i> )	CHB	Hématologie et transfusion
M. Gilles <b>TOURNEL</b>	HCN	Médecine Légale
Mr Olivier <b>TROST</b>	HCN	Anatomie -Chirurgie Maxillo-Faciale
Mr Jean-Jacques <b>TUECH</b>	HCN	Chirurgie digestive
Mr Benoît <b>VEBER</b>	HCN	Anesthésiologie - Réanimation chirurgicale
Mr Pierre <b>VERA</b>	CHB	Biophysique et traitement de l'image
Mr Eric <b>VERIN</b>	Les Herbiers	Médecine Physique et de Réadaptation
Mr Eric <b>VERSPYCK</b>	HCN	Gynécologie obstétrique
Mr Olivier <b>VITTECOQ</b>	HC	Rhumatologie
Mr David <b>WALLON</b>	HCN	Neurologie
Mme Marie-Laure <b>WELTER</b>	HCN	Physiologie

## **MAITRES DE CONFERENCES DES UNIVERSITES – PRATICIENS HOSPITALIERS**

Mme Najate <b>ACHAMRAH</b>	HCN	Nutrition
Mme Elodie <b>ALESSANDRI-GRADT</b>	HCN	Virologie
Mme Noëlle <b>BARBIER-FREBOURG</b>	HCN	Bactériologie – Virologie
Mr Emmanuel <b>BESNIER</b>	HCN	Anesthésiologie - Réanimation
Mme Carole <b>BRASSE LAGNEL</b>	HCN	Biochimie
Mme Valérie <b>BRIDOUX HUYBRECHTS</b>	HCN	Chirurgie Vasculaire
Mr Gérard <b>BUCHONNET</b>	HCN	Hématologie
Mme Mireille <b>CASTANET</b>	HCN	Pédiatrie
Mme Nathalie <b>CHASTAN</b>	HCN	Neurophysiologie
M. Vianney <b>GILARD</b>	HCN	Neurochirurgie
Mr Serge <b>JACQUOT</b>	UFR	Immunologie
Mr Joël <b>LADNER</b>	HCN	Épidémiologie, économie de la santé
Mr Jean-Baptiste <b>LATOUCHE</b>	UFR	Biologie cellulaire
M. Florent <b>MARGUET</b>	HCN	Histologie
Mme Chloé <b>MELCHIOR</b>	HCN	Gastroentérologie
M. Sébastien <b>MIRANDA</b>	HCN	Chirurgie Vasculaire
Mr Thomas <b>MOUREZ</b> ( <i>détachement</i> )	HCN	Virologie
Mr Gaël <b>NICOLAS</b>	UFR	Génétique
Mme Muriel <b>QUILLARD</b>	HCN	Biochimie et biologie moléculaire
Mme Laëtitia <b>ROLLIN</b>	HCN	Médecine du Travail
Mme Pascale <b>SAUGIER-VEBER</b>	HCN	Génétique
M. Abdellah <b>TEBANI</b>	HCN	Biochimie et Biologie Moléculaire
Mme Anne-Claire <b>TOBENAS-DUJARDIN</b>	HCN	Anatomie
Mr Julien <b>WILS</b>	HCN	Pharmacologie

## **PROFESSEUR AGREGÉ OU CERTIFIÉ**

Mr Thierry <b>WABLE</b>	UFR	Communication
Mme Mélanie <b>AUVRAY-HAMEL</b>	UFR	Anglais

## **ATTACHE TEMPORAIRES D'ENSEIGNEMENT ET DE RECHERCHE à MI-TEMPS**

Mme Justine <b>SAULNIER</b>	UFR	Biologie
-----------------------------	-----	----------

## II - PHARMACIE

### PROFESSEURS DES UNIVERSITES

Mr Jérémy <b>BELLIEN</b> (PU-PH)	Pharmacologie
Mr Thierry <b>BESSON</b>	Chimie Thérapeutique
Mr Jean <b>COSTENTIN</b> (Professeur émérite)	Pharmacologie
Mme Isabelle <b>DUBUS</b>	Biochimie
Mr Abdelhakim <b>EL OMRI</b>	Pharmacognosie
Mr François <b>ESTOUR</b>	Chimie Organique
Mr Loïc <b>FAVENNEC</b> (PU-PH)	Parasitologie
Mr Jean Pierre <b>GOULLE</b> (Professeur émérite)	Toxicologie
Mme Christelle <b>MONTEIL</b>	Toxicologie
Mme Martine <b>PESTEL-CARON</b> (PU-PH)	Microbiologie
Mr Rémi <b>VARIN</b> (PU-PH)	Pharmacie clinique
Mr Jean-Marie <b>VAUGEOIS</b>	Pharmacologie
Mr Philippe <b>VERITE</b>	Chimie analytique

### MAITRES DE CONFERENCES DES UNIVERSITES

Mme Cécile <b>BARBOT</b>	Chimie Générale et Minérale
Mr Frédéric <b>BOUNOURE</b>	Pharmacie Galénique
Mr Thomas <b>CASTANHEIRO MATIAS</b>	Chimie Organique
Mr Abdeslam <b>CHAGRAOUI</b>	Physiologie
Mme Camille <b>CHARBONNIER (LE CLEZIO)</b>	Statistiques
Mme Elizabeth <b>CHOSSON</b>	Botanique
Mme Marie Catherine <b>CONCE-CHEMTOB</b>	Législation pharmaceutique et économie de la santé
Mme Cécile <b>CORBIERE</b>	Biochimie
Mme Nathalie <b>DOURMAP</b>	Pharmacologie
Mme Isabelle <b>DUBUC</b>	Pharmacologie
Mme Dominique <b>DUTERTE- BOUCHER</b>	Pharmacologie
Mr Gilles <b>GARGALA</b> (MCU-PH)	Parasitologie
Mme Nejla <b>EL GHARBI-HAMZA</b>	Chimie analytique
Mme Marie-Laure <b>GROULT</b>	Botanique
Mr Chervin <b>HASSEL</b>	Biochimie et Biologie Moléculaire

Mme Maryline <b>LECOINTRE</b>	Physiologie
Mme Hong <b>LU</b>	Biologie
Mme Marine <b>MALLETER</b>	Toxicologie
M. Jérémie <b>MARTINET</b> (MCU-PH)	Immunologie
M. Romy <b>RAZAKANDRAINIBÉ</b>	Parasitologie
Mme Tiphaine <b>ROGEZ-FLORENT</b>	Chimie analytique
Mr Mohamed <b>SKIBA</b>	Pharmacie galénique
Mme Malika <b>SKIBA</b>	Pharmacie galénique
Mme Christine <b>THARASSE</b>	Chimie thérapeutique
Mr Frédéric <b>ZIEGLER</b>	Biochimie

#### **PROFESSEURS ASSOCIES**

Mme Cécile <b>GUERARD-DETUNCQ</b>	Pharmacie officinale
Mme Caroline <b>BERTOUX</b>	Pharmacie

#### **PAU-PH**

M. Mikaël **DAOUPHARS**

#### **PROFESSEUR CERTIFIE**

Mme Mathilde <b>GUERIN</b>	Anglais
----------------------------	---------

#### **ASSISTANTS HOSPITALO-UNIVERSITAIRES**

Mme Alice <b>MOISAN</b>	Virologie
M. Henri <b>GONDÉ</b>	Pharmacie

#### **ATTACHES TEMPORAIRES D'ENSEIGNEMENT ET DE RECHERCHE**

Mme Soukaina <b>GUAOUA-ELJADDI</b>	Informatique
Mme Clémence <b>MEAUSOONE</b>	Toxicologie

#### **ATTACHE TEMPORAIRE D'ENSEIGNEMENT**

Mme Ramla <b>SALHI</b>	Pharmacognosie
------------------------	----------------

<b>LISTE DES RESPONSABLES DES DISCIPLINES PHARMACEUTIQUES</b>
---

Mme Cécile <b>BARBOT</b>	Chimie Générale et minérale
Mr Thierry <b>BESSON</b>	Chimie thérapeutique
Mr Abdeslam <b>CHAGRAOUI</b>	Physiologie
Mme Elisabeth <b>CHOSSON</b>	Botanique
Mme Marie-Catherine <b>CONCE-CHEMTOB</b>	Législation et économie de la santé
Mme Isabelle <b>DUBUS</b>	Biochimie
Mr Abdelhakim <b>EL OMRI</b>	Pharmacognosie
Mr François <b>ESTOUR</b>	Chimie organique
Mr Loïc <b>FAVENNEC</b>	Parasitologie
Mr Michel <b>GUERBET</b>	Toxicologie
Mme Martine <b>PESTEL-CARON</b>	Microbiologie
Mr Mohamed <b>SKIBA</b>	Pharmacie galénique
Mr Rémi <b>VARIN</b>	Pharmacie clinique
M. Jean-Marie <b>VAUGEOIS</b>	Pharmacologie
Mr Philippe <b>VERITE</b>	Chimie analytique



## ENSEIGNANTS MONO-APPARTENANTS

### PROFESSEURS

Mr Paul <b>MULDER</b> (phar)	Sciences du Médicament
Mme Su <b>RUAN</b> (med)	Génie Informatique

### MAITRES DE CONFERENCES

Mr Sahil <b>ADRIOUCH</b> (med)	Biochimie et biologie moléculaire (Unité Inserm 905)
Mme Gaëlle <b>BOUGEARD-DENOYELLE</b> (med)	Biochimie et biologie moléculaire (UMR 1079)
Mme Carine <b>CLEREN</b> (med)	Neurosciences (Néovasc)
M. Sylvain <b>FRAINEAU</b> (med)	Physiologie (Inserm U 1096)
Mme Pascaline <b>GAILDRAT</b> (med)	Génétique moléculaire humaine (UMR 1079)
Mr Nicolas <b>GUEROUT</b> (med)	Chirurgie Expérimentale
Mme Rachel <b>LETELLIER</b> (med)	Physiologie
Mr Antoine <b>OUVRARD-PASCAUD</b> (med)	Physiologie (Unité Inserm 1076)
Mr Frédéric <b>PASQUET</b>	Sciences du langage, orthophonie
Mme Anne-Sophie <b>PEZZINO</b>	Orthophonie
Mme Christine <b>RONDANINO</b> (med)	Physiologie de la reproduction
Mr Youssan Var <b>TAN</b>	Immunologie
Mme Isabelle <b>TOURNIER</b> (med)	Biochimie (UMR 1079)

### **DIRECTEUR ADMINISTRATIF : M. Jean-Sébastien VALET**

*HCN - Hôpital Charles Nicolle*

*HB - Hôpital de BOIS GUILLAUME*

*CB - Centre Henri Becquerel*

*CHS - Centre Hospitalier Spécialisé du Rouvray*

*CRMPR - Centre Régional de Médecine Physique et de Réadaptation*

*SJ – Saint Julien Rouen*

Par délibération en date du 3 mars 1967, la faculté a arrêté que les opinions émises dans les dissertations qui lui seront présentées doivent être considérées comme propres à leurs auteurs et qu'elle n'entend leur donner aucune approbation ni improbation.

*« Savoir s'étonner à propos est le premier pas fait sur la route de la découverte. »*

Louis Pasteur

## **REMERCIEMENTS**

### **AUX MEMBRES DU JURY**

**À Monsieur le Professeur David MALTÊTE,**

Pour me faire l'honneur de présider cette thèse. Je vous remercie également pour votre disponibilité et votre enthousiasme à partager votre expertise des mouvements anormaux. Veuillez recevoir l'expression de toute ma reconnaissance.

**À Monsieur le Docteur Bertrand BOURRE,** mon directeur de thèse, guide et mentor

Pour ta confiance accordée pour réaliser ce travail, ton encadrement avisé et ton soutien solide. Merci aussi pour ton engagement dans ma formation dans le domaine de l'inflammatoire. Reçois toute ma gratitude.

**À Monsieur le Professeur Emmanuel GERARDIN,**

Vous me faites l'honneur de juger cette thèse. Soyez remercié pour votre engagement. Recevez l'expression de mon profond respect.

**À Madame le Docteur Lou GRANGEON,**

Pour avoir accepté de juger ce travail. Pour ta rigueur modèle, ta motivation à enseigner et ta bienveillance. Je t'adresse mes sincères remerciements.

### **AUX AUTRES PARTICIPANTS À CE TRAVAIL**

**À l'équipe de neuroradiologie,** pour l'acquisition et l'interprétation des IRM.

**À l'équipe de PIXYL, Benjamin LAMBERT, Maxime LOUIS, Thibaud POUCH et Alan TUCHOLKA.** Merci pour votre aide et votre investissement dans cette étude. Je vous souhaite le meilleur dans le développement de PIXYL.

**À Thibaut PRESSAT LAFFOUILHERE,** pour sa supervision éclairée des statistiques et nos échanges qui ont élargi ma culture statistique.

**Aux patients** qui ont accepté de participer au registre EDMUS et permis ce travail de recherche exploratoire. **À Christine VIMONT** qui gère la base rouennaise, pour son aide.

## **À MES AMIS ET COLLÈGUES**

Pour votre soutien infaillible et votre contribution inestimable à ma formation.

Adèle, Adina, Adrien, Agathe, Alexandra, Alexandre, Alexis, Alice, Aline, Anaïs, Anne-Claire, Anne-Laure, Anne-Marie, Antoine, Aude, Audrey, Axel, Benjamin, Bertrand, Billal, Brigitte, Bruno, Camille, Catherine, Cécile, Christelle, Christine, Claire, Claire-Sophie, Clara, Clémence, Clément, Coline, Constance, Corinne, Damien, Daniel, Daphné, David, Delphine, Didier, Élisabeth, Élise, Elodie, Estelle, Evelyne, Fabien, Fanny, Félix, Flora, Florian, Floriane, François-Xavier, Gabriel, Gabriella, Gaël, Gaëtan, Gérald, Giovanni, Grégory, Guillaume, Gülden, Hervé, Isabelle, Jacques, Jad, James-William, Jasmine, Jean, Jean-Paul, Jean-Philippe, Jérôme, Jonathan, Julie, Justine, Laura, Laurent, Léa, Léna, Linda, Lotfi, Lou, Lucie, Lucile, Margaux, Marie, Marie-Laure, Mathieu, Mathilde, Maxime, Maximilien, Mélina, Michel, Michèle, Morgane, Natacha, Nathalie, Nicolas, Olivier, Omar, Ozlem, Pascale, Patrick, Paul-Alexandre, Pauline, Philippe, Pierre, Raphaël, Roland, Romain, Salimata, Simon, Sofiane, Sophie, Stéphane, Sydney, Thibault, Thierry, Thomas, Trésor, Valérie, Vianney, Victor, Vincent, Yannick, Ygal, Yvan, Zélie...

**À l'équipe de Neurologie et Neurophysiologie du CHU de Rouen** pour votre accompagnement de premier plan. Une attention particulière pour David W qui a sauvé mon année recherche.

**À l'équipe de Neurologie du CH d'Évreux** pour votre accueil chaleureux.

**À l'équipe de Neurologie du GH du Havre** pour votre accueil chaleureux.

**À l'équipe de Médecine Interne du CHU de Rouen** pour votre accueil chaleureux.

**À l'équipe de Neurochirurgie du CHU de Rouen** pour votre accueil chaleureux.

**À l'équipe de Génétique du CHU de Rouen** pour votre accueil chaleureux.

**Aux neurologues du CH d'Elbeuf et aux neurologues libéraux** pour les bons moments partagés ensemble.

**Aux enseignants** qui m'ont formé, **À tous ceux** qui m'ont soutenu et encouragé à publier.

**À l'équipe de l'ICM** qui m'a initié à la recherche en génétique moléculaire.

**À mes co-internes, Aux personnes rencontrées dans les autres équipes, tous ceux** qui ont partagé ma formation.

**À mes meilleurs amis, musiciens ou non.**

## **À MA FAMILLE ET MES PROCHES**

Pour votre affection et votre support.

## TABLE DES MATIÈRES

Liste des abréviations.....	15
Introduction.....	16
I. Définition, physiopathologie et clinique de la SEP .....	16
II. IRM .....	19
III. Diagnostic .....	20
IV. Évaluation et suivi des traitements .....	22
V. Interprétation et méthodes automatiques .....	23
VI. Objectifs de l'étude .....	27
Matériels et méthodes .....	28
Résultats.....	31
I. Caractéristiques de la population étudiée .....	31
II. Marqueurs radiologiques FUSION et PIXYL.....	34
III. Comparaison des performances de la mesure d'activité radiologique FUSION et PIXYL pour identifier l'échec thérapeutique .....	41
IV. Recherche de seuils lésionnels PIXYL optimaux.....	48
V. Combinaison de marqueurs radiologiques dans des modèles de régression logistique d'identification de l'échec thérapeutique global.....	51
VI. Analyses de concordance des marqueurs PIXYL avec la FUSION et l'échec thérapeutique global.....	57
VII. Vitesse de survenue de l'échec thérapeutique. ....	61
VIII. Masques de segmentation PIXYL.....	63
Discussion.....	66
Conclusion et perspectives.....	77
Bibliographie .....	79
Annexes .....	83
Summary.....	93
Résumé.....	94

## LISTE DES ABRÉVIATIONS

CIS : Syndrome Cliniquement Isolé

DIT : Dissémination Temporelle

DIS : Dissémination Spatiale

EDMUS : European Database for Multiple Sclerosis

EDSS : Expanded Disability Status Scale

FUSION : fusion de séquences FLAIR pour comparaison entre 2 IRM

IRM : Imagerie par Résonance Magnétique

iqr : intervalle interquartile

MAGNIMS : Magnetic Resonance Imaging in Multiple Sclerosis

NEDA : No Evidence of Disease Activity

OFSEP : Observatoire Français de la Sclérose en Plaques

PP : Primaire Progressive

RR : Rémittente Récurrente

SEP : Sclérose En Plaques

SF : Score Fonctionnel

SNC : Système Nerveux Central

SP : Secondairement Progressive

## **INTRODUCTION**

### **I. Définition, physiopathologie et clinique de la SEP**

La sclérose en plaques (SEP) est une maladie inflammatoire auto-immune chronique du système nerveux central (SNC) (9). Sa prévalence est estimée à 156 pour 100 000 habitants en France soit plus de 90 000 personnes (16). Il s'agit de la première cause de handicap sévère non traumatique chez le sujet jeune, l'âge moyen d'apparition étant de 30 ans. Une prédominance féminine est observée avec un sex-ratio à 3:1.

La physiopathologie de la SEP est encore imparfaitement comprise et repose sur le passage de lymphocytes auto-réactifs à travers la barrière hémato-encéphalique (7). Un déséquilibre entre les processus d'activation et de régulation immunitaires est provoqué par ces lymphocytes auto-réactifs contre le SNC. Il entraîne une inflammation responsable de démyélinisation en plaques et de neurodégénérescence. Il est historiquement décrit deux formes évolutives de SEP, rémittente-récurrente (RR) dans 90 % des cas et progressive dans 10 % (primaire et secondairement progressives (PP et SP)) (32). Cette description repose sur la prépondérance de poussées démyélinisantes aiguës dans la forme RR versus une progression continue du handicap en l'absence de poussées dans les formes progressives. La nouvelle classification de Lublin complète la description évolutive de la SEP par les notions d'activité (active ou non active) et de progression (avec ou sans progression pour les formes progressives) rendant compte de la coexistence des deux processus quelle que soit la forme clinique (33). L'activité est définie par la présence de poussées cliniques ou d'évolution IRM (lésions rehaussées par le gadolinium, nouvelles lésions T2 ou ayant augmenté de taille). La progression correspond à l'aggravation progressive confirmée du handicap neurologique sur le plan clinique et il n'existe pas encore de consensus pour définir la progression radiologique. Cette classification phénotypique de la SEP proposée par Lublin est illustrée par la Figure 1 (32,33).

Les manifestations cliniques sont la résultante des lésions encéphaliques et / ou médullaires. Leur expression dépend de leur localisation et de leur taille lorsqu'elles touchent des zones fonctionnelles (9). Les plaques de démyélinisation, qui entraînent des troubles de conduction axonale (ralentissement et décharges spontanées), et la mort neuronale sollicitent la réserve cérébrale qui compense de manière limitée leur retentissement clinique selon la taille et la localisation des lésions. La sévérité clinique peut être évaluée à l'aide du score EDSS (Expanded Disability Status Scale) étendu de 0.0 à 10.0

(décès) (Tableau 1) (27). Ce score est historiquement l’outil le plus utilisé au quotidien et en recherche. La sévérité du handicap cotée dans l’EDSS est précisée par 8 scores fonctionnels (SF) portant sur les fonctions pyramidale, cérébelleuse, du tronc cérébral, de la sensibilité, sphinctériennes, visuelle, cérébrale et autres (Annexe 1). L’évolutivité clinique est définie par la survenue d’une poussée ou l’aggravation de l’EDSS. Une poussée est définie comme l’apparition ou l’aggravation subaiguë en quelques heures à quelques jours de symptômes neurologiques durant au moins 24h avant une récupération progressive plus ou moins complète, et non expliquée par une fluctuation de l’état clinique liée notamment à de la fièvre, une infection ou de la fatigue. Une aggravation de l’EDSS est reconnue significative lors du passage de 0 à 1.5, de la majoration d’au moins 1 point pour un EDSS antérieur non nul et < 6 et d’au moins 0,5 point pour un EDSS antérieur ≥ 6.

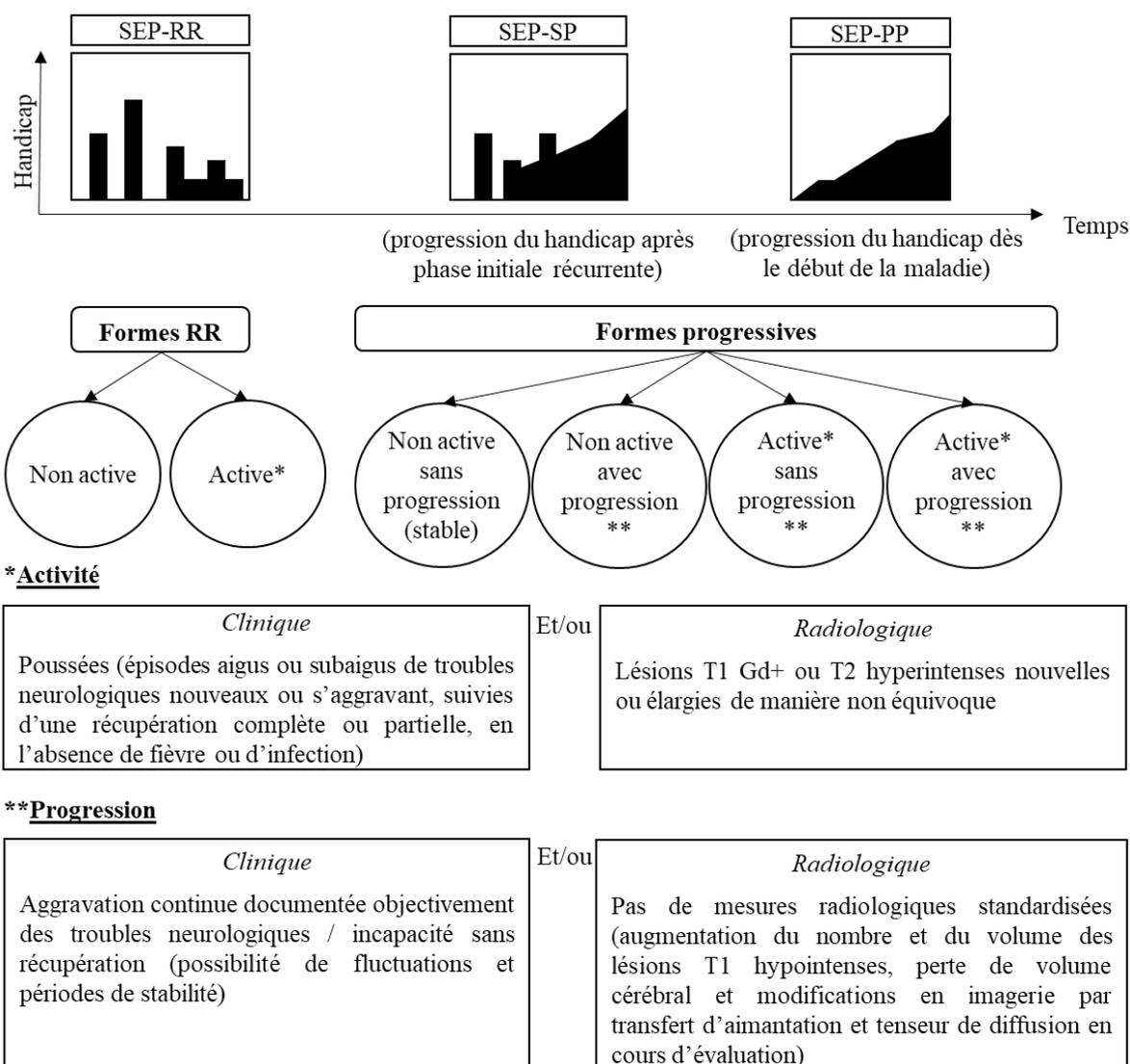


Figure 1. Classification phénotypique de la SEP adaptée de Lublin (32,33)

Score	Critères
0	Examen neurologique normal (tous systèmes fonctionnels (SF) à 0; SF 1 mental acceptable).
1.0	Absence de handicap fonctionnel, signes minimales d'atteinte d'une des fonctions (SF 1, à l'exclusion du SF mental).
1.5	Absence de handicap fonctionnel, signes minimales dans plus d'un SF (plus d'un SF 1, à l'exclusion du SF mental).
2.0	Handicap minimal d'un des SF (1 SF 2, les autres 0 ou 1).
2.5	Handicap minimal dans 2 SF (2 SF 2, les autres 0 ou 1).
3.0	Handicap modéré dans un SF (1 SF score 3, les autres 0 ou 1) ; ou atteinte minimale de 3 ou 4 fonctions (3 ou 4 SF 2 ; les autres 0 ou 1), mais malade totalement ambulatoire.
3.5	Totalement ambulatoire, mais atteinte modérée dans un SF (SF 3) et 1 ou 2 SF 2; ou 2 SF 3 ; ou 5 SF 2 (les autres 0 ou 1).
4.0	Malade totalement autonome pour la marche, vaquant à ses occupations 12h par jour malgré une gêne fonctionnelle relativement importante : 1 SF à 4 (les autres 0 ou 1), ou association de niveaux inférieurs dépassant les limites des degrés précédents. Capable de marcher 500 m environ sans aide ni repos.
4.5	Malade autonome pour la marche, vaquant à ses occupations la majeure partie de la journée, capable de travailler une journée entière, mais pouvant parfois être limité dans ses activités ou avoir besoin d'une aide minimale, handicap relativement sévère : un SF 4 (les autres 0 ou 1), ou association de niveaux inférieurs dépassant les limites des degrés précédents. Capable de marcher 300m environ sans aide ni repos.
5.0	Capable de marcher environ 200 m sans aide ni repos, handicap suffisamment sévère pour entraver l'activité d'une journée normale. (En général un SF 5, les autres 0 ou 1, ou association de niveaux plus faibles dépassant ceux du grade 4.0).
5.5	Capable de marcher environ 100 m sans aide ni repos ; handicap suffisamment sévère pour empêcher l'activité d'une journée normale. (En général un SF 5, les autres 0 ou 1, ou association de niveaux plus faibles dépassant ceux du grade 4.0).
6.0	Aide unilatérale (cane, canne anglaise, béquille) constante ou intermittente nécessaire pour parcourir environ 100 m avec ou sans repos intermédiaire. (En général association de SF comprenant plus de 2 SF 3+).
6.5	Aide permanente bilatérale (cannes, cannes anglaises, béquilles) nécessaire pour marcher 20 m sans s'arrêter. (En général association de SF comprenant plus de 2 SF 3+).
7.0	Incapable de marcher plus de 5 m même avec aide ; essentiellement confiné au fauteuil roulant; fait avancer lui-même son fauteuil et effectue le transfert; est au fauteuil roulant au moins 12 h par jour. (En général association de SF comprenant plus d'un SF 4+; très rarement, SF 5 pyramidal seulement).
7.5	Incapable de faire plus de quelques pas; strictement confiné au fauteuil roulant; a parfois besoin d'une aide pour le transfert; peut faire avancer lui-même son fauteuil mais ne peut y rester toute la journée; peut avoir besoin d'un fauteuil électrique. (En général association de SF comprenant plus d'un SF 4+).
8.0	Essentiellement confiné au lit ou au fauteuil, ou promené en fauteuil par une autre personne; peut rester hors du lit la majeure partie de la journée; conserve la plupart des fonctions élémentaires; conserve en général l'usage effectif des bras. (En général SF 4+ dans plusieurs systèmes).
8.5	Confiné au lit la majeure partie de la journée ; garde un usage partiel des bras ; conserve quelques fonctions élémentaires. (En général SF 4+ dans plusieurs systèmes).
9.0	Patient grabataire ; peut communiquer et manger. (En général SF 4+ dans plusieurs systèmes).
9.5	Patient totalement impotent, ne peut plus manger ou avaler, ni communiquer. (En général SF 4+ dans presque tous les systèmes).
10	Décès lié à la SEP.

Tableau 1. Quotation du score EDSS (adapté d'EDMUS.org) (SF : Score Fonctionnel) (27)

## II. IRM

L'imagerie par résonance magnétique (IRM) cérébrale et médullaire permet le diagnostic (cf III. Diagnostic) et le suivi des lésions de la SEP même si seule une partie des lésions est visible (24). C'est un marqueur objectif et plus sensible que la clinique. En effet, pour chaque poussée clinique, il est admis que 10 lésions IRM asymptomatiques sont retrouvées (9). L'activité radiologique est évaluée par l'apparition de nouvelles lésions ou l'extension / élargissement de lésions déjà présentes.

Différentes séquences magnétiques sont utilisées. Des recommandations internationales (Magnetic Resonance Imaging in Multiple Sclerosis, MAGNIMS) et nationales (Observatoire Français de la Sclérose en Plaques, OFSEP) existent afin de rendre les pratiques homogènes, comparables et fiables dans tous les centres (Tableau 2) (2,52). La nécessité d'une injection de gadolinium a été fortement réduite lors de l'actualisation de ces recommandations en 2020.

Il est recommandé de réaliser une IRM cérébrale au diagnostic, lors d'un changement de traitement (avant et dans les 6 mois suivant son introduction, délai minimal pour que le traitement soit pleinement efficace ; cette dernière est dite IRM de rebase) puis lors du suivi sous traitement au moins annuellement ou lors d'une poussée. L'IRM médullaire est recommandée tous les trois ans.

Des spécifications techniques sont également précisées dans ces recommandations. Il s'agit notamment d'un champ magnétique suffisant d'au moins 1,5 Tesla, d'une acquisition veillant à respecter un bon rapport signal / bruit, de préférer les séquences 3D avec des voxels de 1 mm<sup>3</sup> sans espace entre les coupes ou à défaut de coupes 2D fines  $\leq 3$  mm d'épaisseur sans espace entre les coupes.

<b>Protocole IRM OFSEP 2020</b>		
	<b>IRM cérébrale</b>	<b>IRM médullaire</b>
<b>RECOMMANDÉ</b>	<ul style="list-style-type: none"> <li>• <b>Protocole OFSEP standard</b> 3D T1 millimétrique DWI axiale + dADC 3D FLAIR millimétrique</li> <li>• <b>Protocole OFSEP réduit (suivi trimestriel LEMP)</b> DWI axiale + dADC 3D FLAIR millimétrique</li> </ul>	<ul style="list-style-type: none"> <li>• <b>Protocole OFSEP standard</b> T2 sagittale</li> </ul>
<b>OPTIONNEL</b>	3D T1 gadolinium * millimétrique 2D TSE DP / T2 axiale ou 3D T2 DTI ≥ 15 directions 2D T2 EG (pour un 1 <sup>er</sup> diagnostic) SWI 3D DIR	T1 sagittale gadolinium * T2 EG axiale T1 axiale gadolinium * STIR sagittale
<p><b>* Quand injecter du gadolinium (0,1 mmol/kg) ?</b>            L'OFSEP recommande l'utilisation de gadolinium macrocyclique.</p> <ul style="list-style-type: none"> <li>- Au début du suivi (diagnostic, IRM antérieures non disponibles)</li> <li>- À l'initialisation d'un traitement de fond</li> <li>- Six mois après une initialisation de traitement de fond</li> <li>- <i>[Recommandé]</i> En cas de poussée</li> <li>- <i>[Optionnel]</i> Pour rechercher une activité en vue d'une indication thérapeutique</li> </ul>		

Tableau 2. Synthèse du protocole IRM SEP recommandé par l'OFSEP depuis 2020 (2)

### **III. Diagnostic**

Le diagnostic de SEP repose sur la dissémination spatiale (DIS) et temporelle (DIT) de lésions démyélinisantes du SNC d'aspect évocateur en l'absence de diagnostic différentiel. Ces critères diagnostiques ont été révisés à plusieurs reprises dont la dernière fois en 2017 (critères de McDonald 2017, [Tableau 3](#)) (50). L'IRM permet d'objectiver la présence de ces lésions sous forme d'hypersignaux de la substance blanche en séquence T2 / FLAIR. Elles peuvent être rehaussées par l'injection de gadolinium si elles sont récentes (< 3 mois) et apparaître sous forme d'hyposignaux appelés « black holes » (« trous noirs ») en séquence T1 en cas d'atteinte marquée ([Figure 2. Lésions IRM dans la SEP](#)).

Une expertise clinique et radiologique est indispensable pour bien évaluer la maladie du fait de nombreux diagnostics différentiels (9). Les hypersignaux T2 / FLAIR doivent être évalués attentivement afin de repérer les atypies bien illustrées par la revue de Filippi et al. (13). En pratique, une taille seuil d'un diamètre lésionnel  $\geq 3$  mm est recommandée afin de renforcer la spécificité des lésions retenues comme significatives de l'atteinte inflammatoire de la SEP (21,50).

	<u>SEP rémittente récurrente</u> définie par :
DIS	<ul style="list-style-type: none"> <li>• Soit (i) preuve clinique objective de <math>\geq 2</math> lésions ou preuve clinique objective d'une lésion avec preuve historique raisonnable d'une précédente poussée impliquant un autre site du SNC</li> <li>• ou (ii) <math>\geq 1</math> lésion T2 dans au moins 2 des 4 régions du SNC typiques de la SEP (périventriculaire, juxta-corticale, infratentorielle et moelle épinière)</li> </ul>
DIT	<ul style="list-style-type: none"> <li>• Soit (i) <math>\geq 2</math> poussées séparées d'au moins 1 mois</li> <li>• ou (ii) présence simultanée de lésions asymptomatiques rehaussées et non rehaussées par le gadolinium à n'importe quel moment</li> <li>• ou (iii) apparition d'une nouvelle lésion T2 ou réhaussée par le gadolinium lors d'une IRM de suivi</li> <li>• ou (iv) démonstration d'une synthèse intrathécale prouvée par la présence de bandes oligoclonales spécifiques au LCR (Liquide Céphalo-Rachidien)</li> </ul>
	<u>SEP primaire progressive</u> définie par :
DIS	<p>Au moins 2 des 3 critères suivants :</p> <ul style="list-style-type: none"> <li>• <math>\geq 1</math> lésion cérébrale T2 caractéristique (périventriculaire, juxta-corticale, infratentorielle)</li> <li>• <math>\geq 2</math> lésions médullaires</li> <li>• LCR positif (bandes oligoclonales et / ou index IgG élevé)</li> </ul>
DIT	<ul style="list-style-type: none"> <li>• Une année de progression de la maladie (déterminée rétrospectivement ou prospectivement)</li> </ul>

Tableau 3. Critères de MacDonald 2017 pour le diagnostic de SEP (50)

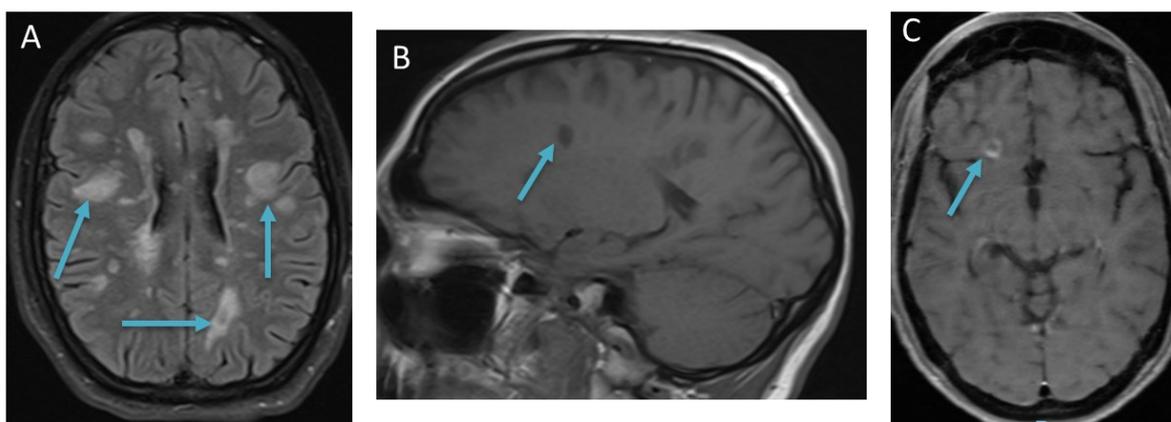


Figure 2. Lésions IRM dans la SEP (A. Hypersignaux FLAIR périventriculaire et juxtacorticaux ; B. Hyposignal T1 ; C. Rehaussement en anneau suite à l'injection de Gadolinium)

#### **IV. Évaluation et suivi des traitements**

Les poussées, l'aggravation du handicap (EDSS) et les nouvelles lésions IRM reflètent l'activité de la maladie (38). La notion de fenêtre d'opportunité thérapeutique définit le fait qu'il est important d'optimiser précocement le traitement pour prévenir des dommages neurologiques irréversibles et réduire le risque de progression de la maladie (Figure 3) (20). L'objectif des traitements est donc de stopper ou du moins freiner l'activité de la maladie en ciblant principalement l'auto-réactivité inflammatoire du système immunitaire dirigée contre la myéline ; la promotion de la remyélinisation et la lutte contre la neurodégénérescence étant des cibles encore mal couvertes. L'IRM est un bon marqueur du contrôle de la maladie par les traitements et prédit l'évolution clinique future (46). L'IRM permet d'évaluer l'efficacité thérapeutique en s'assurant de l'absence de progression de la charge lésionnelle et de l'absence d'évolution en taille des lésions présentes, ces marqueurs radiologiques étant plus sensibles et précoces que l'évolution clinique.

L'absence d'efficacité du traitement doit faire envisager une modification vers une molécule plus efficace selon une stratégie d'escalade thérapeutique. Il n'existe pas de consensus valable pour toutes les molécules par rapport à un nombre seuil de lésions évolutives pour conclure à une inefficacité thérapeutique, les traitements ne pouvant pas toujours remplir l'objectif idéal d'une maladie sans signes d'activité (NEDA, no evidence of disease activity) (39). Néanmoins, le seuil pertinent de 3 nouvelles lésions est souvent utilisé dans la pratique clinique et repose sur les études de Rio et Sormani pour les patients traités par interféron (Critères de Sormani) (47). Le contrôle clinico-radiologique de la

maladie défini par le NEDA-3 se réfère à l'absence de poussée, de progression de l'EDSS et de lésion évolutive (nouvelle ou élargie) en IRM (39). Le NEDA-4 est encore plus strict et inclut en plus l'absence de perte significative du volume cérébral (atrophie) pour prendre en compte la composante neurodégénérative de la SEP (39).

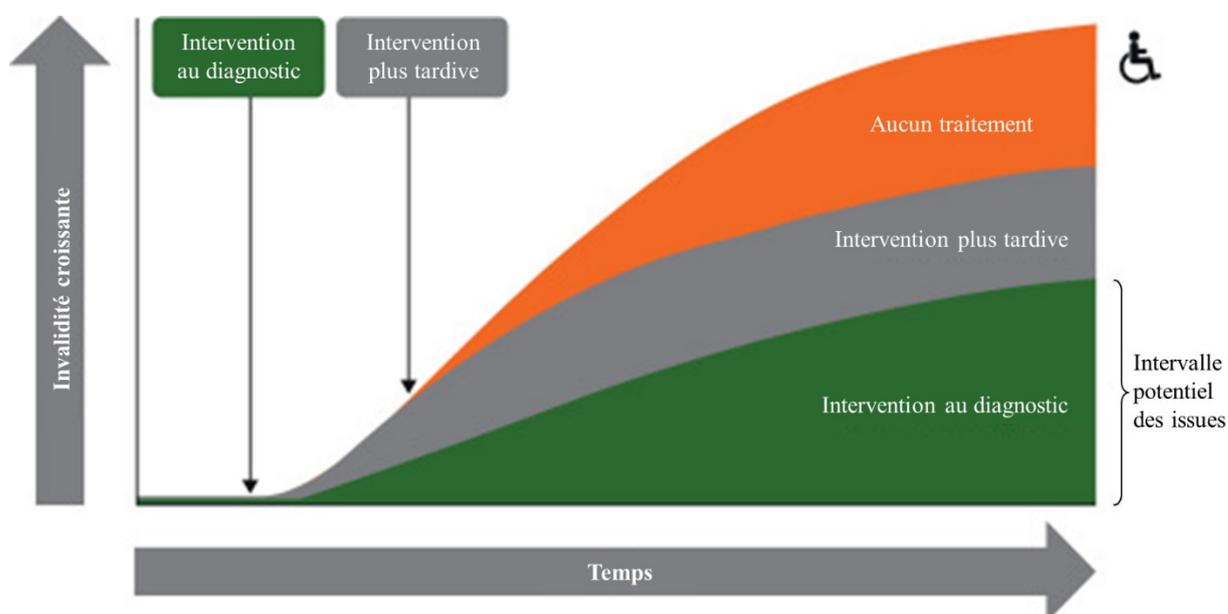


Figure 3. Fenêtre d'opportunité thérapeutique et importance d'une efficacité rapide (adapté de Giovannoni et al., 2016 (20) ; figure originale © 2015 Oxford PharmaGenesis Ltd)

## V. Interprétation et méthodes automatiques

L'interprétation d'une IRM cérébrale demande de la rigueur et du temps. Il est d'autant plus difficile de détecter une progression de la charge lésionnelle qu'elle est déjà importante sur l'IRM de référence et que les nouvelles lésions sont petites. Le suivi précis des lésions démyélinisantes reposerait sur leur segmentation afin de mesurer leur nombre et leur volume. Néanmoins ces mesures étant chronophages, estimées à 60 minutes par imagerie pour un expert (1), elles ne peuvent être appliquées en routine. Le développement de méthodes d'analyse semi-automatiques et automatiques devrait permettre un gain de temps et de précision tout en étant plus objectif et reproductible.

En pratique courante actuelle, deux grandes méthodes semi-automatiques permettent d'améliorer la détection des lésions évolutives par fusion ou soustraction des images 3D recalées (17,37). Elles nécessitent de comparer des images acquises selon les mêmes

paramètres et sur la même machine mais restent sensibles aux difficultés de recalage et certains artefacts (flux sanguin et LCR, modification du volume et de l'intensité du signal IRM).

La méthode de fusion, désignée par le terme FUSION par la suite, est utilisée au CHU de Rouen. En effet, ce post-traitement est proposé par tous les fabricants d'IRM et ne prend que quelques minutes de traitement des images acquises. Les nouvelles lésions apparaissent de manière distincte selon les filtres de couleurs appliqués, en blanc par rapport aux lésions déjà présentes ayant la couleur orange attribuée à l'IRM de référence (Figure 4).

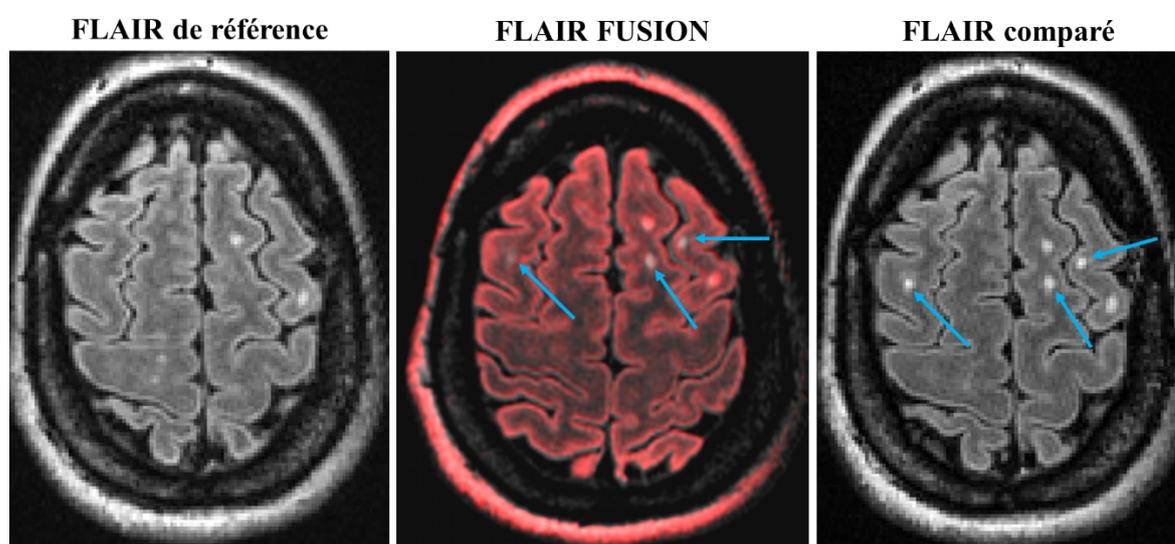


Figure 4. Comparaison de 2 coupes axiales d'IRM cérébrale avec la fusion d'images FLAIR. Apparition de trois nouvelles lésions FLAIR sous-corticales (flèches bleues) ressortant en blanc-orangé sur la fusion des 2 coupes au milieu par rapport aux lésions antérieures en orangé.

Les méthodes de segmentation complètement automatiques sont encore en cours de développement en raison de performances insuffisamment validées en pratique clinique. Elles requièrent des images de la meilleure qualité possible (coupes fines, bon rapport signal / bruit et limitation des artefacts) et un pré-traitement pour normaliser les coupes à analyser (correction du biais de champ magnétique et alignement de référence). De nombreuses méthodes de traitement d'image sont utilisées. Il existe une approche dite supervisée, c'est-à-dire développée à partir d'une base de données d'apprentissage composée de plusieurs segmentations de référence, et une approche non supervisée reposant sur des méthodes de regroupement (18).

La plateforme PIXYL (Pixyl.Neuro.MS platform) est un outil de traitement automatique d'IRM cérébrale développé par une équipe française (56). PIXYL fait appel à de l'intelligence artificielle avec du Deep learning, qui consiste à entraîner un réseau de neurones à partir d'une grande base de données d'IRM manuellement annotées. Il s'agit d'une approche hybride combinant des méthodes supervisées (réseaux de neurones 3D entraînés sur des jeux de données artificiellement augmentés ou agrandis pour améliorer la robustesse du modèle) et non supervisées (champs aléatoires de Markov notamment) (15). La méthodologie exacte et l'ensemble du processus de traitement d'image ne sont pas publics en raison d'une utilisation commerciale avec l'obtention d'un marquage CE. PIXYL réalise la segmentation des lésions encéphaliques de la SEP et la quantification de leur dissémination dans le temps et l'espace. Les données fournies sont :

- Le compte et le volume des lésions T2 / FLAIR par région d'intérêt et caractère évolutif en cas de comparaison de deux imageries (stable, élargie ou nouvelle) ;
- Le compte des lésions rehaussées T1Gd+ ;
- Le compte et le volume des black-holes ou hyposignaux T1 par région d'intérêt et caractère évolutif ;
- La mesure de l'atrophie cérébrale comparée aux normes pour l'âge.

La comparaison de deux imageries au cours du suivi d'un patient permet donc de chiffrer précisément une éventuelle progression radiologique des lésions nouvelles ou élargies. Une synthèse est générée et les masques de segmentation sont visualisables sur une interface en ligne ou directement sur la console d'imagerie où l'outil peut être implémenté (Figure 5). Cette comparaison de deux imageries segmentées fait également appel à un traitement d'images par machine learning et réseaux neuronaux.

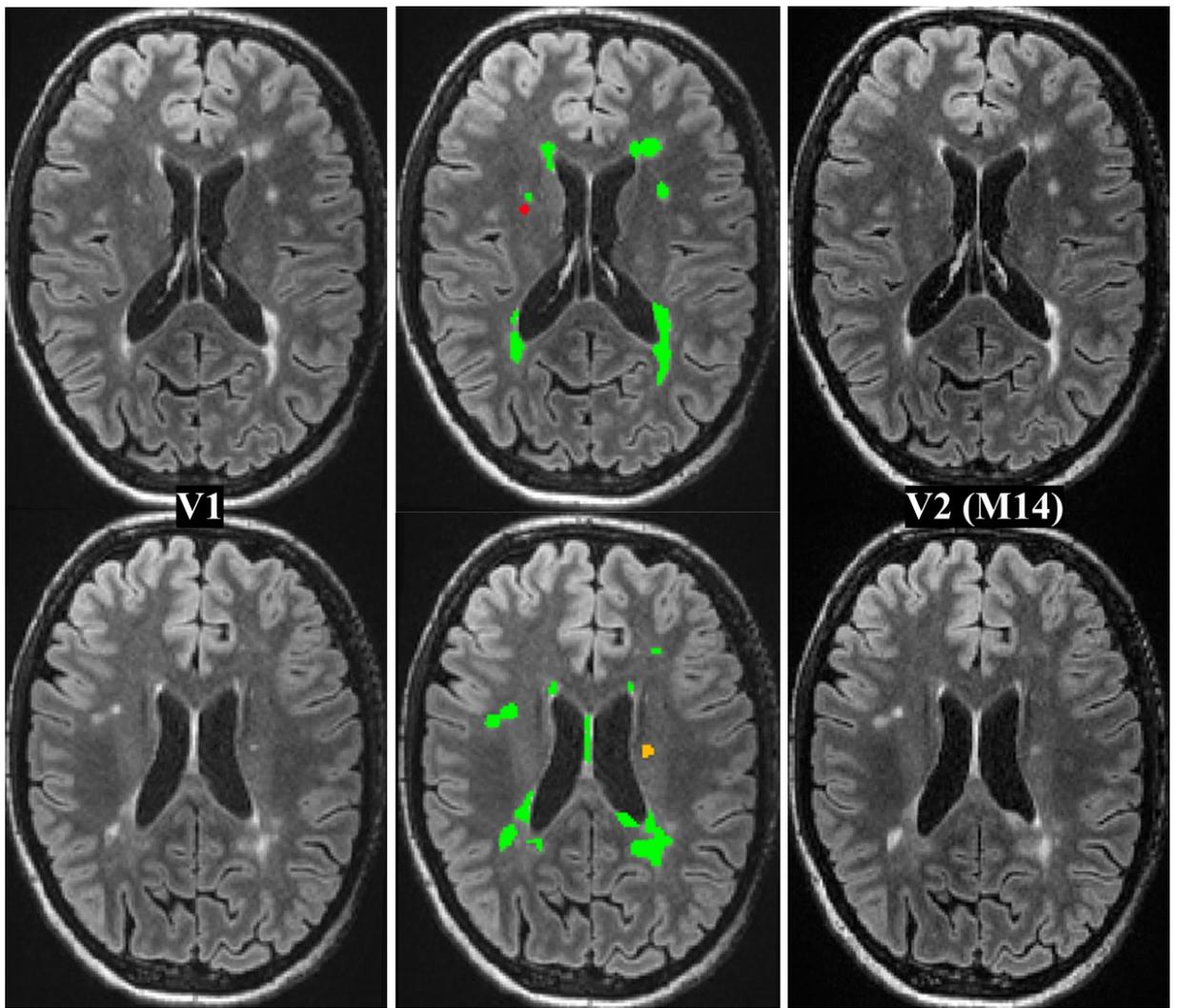


Figure 5. Visualisation des masques de détection de lésions FLAIR entre 2 IRM de suivi par PIXYL. Visite 1 V1 (à gauche) et Visite 2 V2 14 mois plus tard (à droite). Dans le panneau central, les masques de segmentation montrent en vert, les lésions stables ; en rouge, une nouvelle lésion et en orange, une lésion élargie.

## **VI. Objectifs de l'étude**

Compte tenu de l'importance à détecter rapidement l'apparition de nouvelles lésions IRM dans la SEP, nous avons souhaité comparer l'intérêt d'une analyse assistée par segmentation automatique d'IRM cérébrale par la plateforme PIXYL par rapport à l'interprétation radiologique humaine courante assistée par le processus de fusion comparative d'images.

L'objectif principal était de comparer la performance de classification des patients ayant eu un **échec thérapeutique** (défini infra dans les Matériels et méthodes) à un an de l'IRM cérébrale de suivi. Ce délai d'un an correspond en général à l'intervalle régulier de réalisation des IRM cérébrales au cours du suivi de la maladie.

L'objectif principal permettait donc de comparer la pertinence de l'application clinique des deux méthodes pour identifier les patients dont la maladie reste active sous traitement qui présenteront un échec thérapeutique dans l'année suivant l'IRM de suivi.

Les objectifs secondaires étaient de :

- rapporter les mesures de segmentation automatique par PIXYL dans le cadre du suivi courant d'une population SEP variée ;
- préciser les performances de classification de l'activité radiologique des deux méthodes FUSION et PIXYL à la fois pour identifier et prédire la survenue d'un échec thérapeutique respectivement dans les sous-groupes de patients ayant eu un échec entre ou après les deux IRM ;
- comparer la vitesse de survenue d'un échec thérapeutique en fonction de l'activité radiologique mesurée selon les deux méthodes dans le sous-groupe de patients ayant eu un échec thérapeutique après l'IRM de FUSION.

## **MATÉRIELS ET MÉTHODES**

Il s'agissait d'une étude monocentrique rétrospective sur les IRM cérébrales de patients suivis au CHU de Rouen.

### **1. Patients et IRM cérébrales**

Les patients éligibles devaient être suivis au CHU de Rouen (sur les plans clinique et radiologique), avoir eu au moins une IRM cérébrale avec fusion de séquences FLAIR pour comparaison à l'IRM cérébrale antérieure et être inclus dans la cohorte EDMUS (European Database for Multiple Sclerosis) permettant un suivi clinique standardisé prospectif.

Les caractéristiques de la population (âge, sexe, durée d'évolution de la maladie, traitements reçus, handicap mesuré par l'EDSS au moment des IRM puis au cours du suivi ultérieur et survenue de poussées) ont été récupérées à partir de la base EDMUS et d'un retour aux dossiers en cas de données insuffisantes dans EDMUS. Les résultats des IRM cérébrales ont été récupérés à partir des comptes-rendus radiologiques.

L'IRM cérébrale de fusion la plus récente ainsi que l'IRM de référence comparative ont été récupérées puis anonymisées et envoyées à la société PIXYL pour l'analyse sur leur plateforme de segmentation automatique PIXYL.MS version 1.8.1. Les données récupérées pour l'étude étaient la localisation par territoire (infratentorial, profond, périventriculaire et juxta-cortical) et l'identification des lésions FLAIR avec la mesure de leurs dimensions, volume et caractère évolutif (stable, nouvelle ou élargie) sur l'IRM de fusion par rapport à l'IRM de référence. Ces données ont permis de calculer le volume lésionnel total et la somme du nombre de nouvelles lésions FLAIR selon leur caractère évolutif. Les masques de segmentation des lésions FLAIR identifiées ont été visualisés sur les coupes IRM initiales via l'outil ITK-SNAP 3.8.0 (54) ([Figure 5](#) présentée dans l'introduction). Les IRM cérébrales ont été acquises sur trois machines différentes (1,5 et 3 T).

### **2. Critères de jugement**

L'**échec thérapeutique** a été défini par rapport à l'évaluation initiale (lors de l'IRM cérébrale de référence) comme la survenue d'un des événements suivants : i) une poussée, ii) progression de l'EDSS confirmée par deux évaluations : d'au moins 1,5 point si EDSS = 0 au

moment de l'IRM de référence ; d'au moins 1 point si EDSS > 0 et < 6 ou d'au moins 0,5 point si EDSS  $\geq$  6 ou iii) changement de traitement de fond pour inefficacité. Un échec thérapeutique était considéré postérieur à l'IRM de FUSION s'il survenait plus de 30 jours après, temps nécessaire pour que le clinicien puisse prendre connaissance du résultat et qu'il réévalue le patient.

L'indicateur principal retenu pour la comparaison entre les 2 méthodes d'interprétation d'IRM cérébrale était l'aire sous la courbe sensibilité / spécificité (AUC). Les critères secondaires étaient les autres indicateurs statistiques de classification tels que les sensibilités, spécificités et exactitudes. Deux sous-groupes de patients ont été distingués pour préciser des analyses, celui avec survenue d'un échec thérapeutique entre les deux IRM et celui avec survenue d'un échec thérapeutique après l'IRM de FUSION. Les patients ayant eu un deuxième échec thérapeutique pendant la période d'observation pouvaient donc être analysés dans ces deux sous-groupes. Des analyses de survie du délai de survenue de l'échec thérapeutique après l'IRM de FUSION ont pu être réalisées dans le deuxième sous-groupe présenté. La chronologie des différents marqueurs de l'étude a été représentée dans la [Figure 6](#).

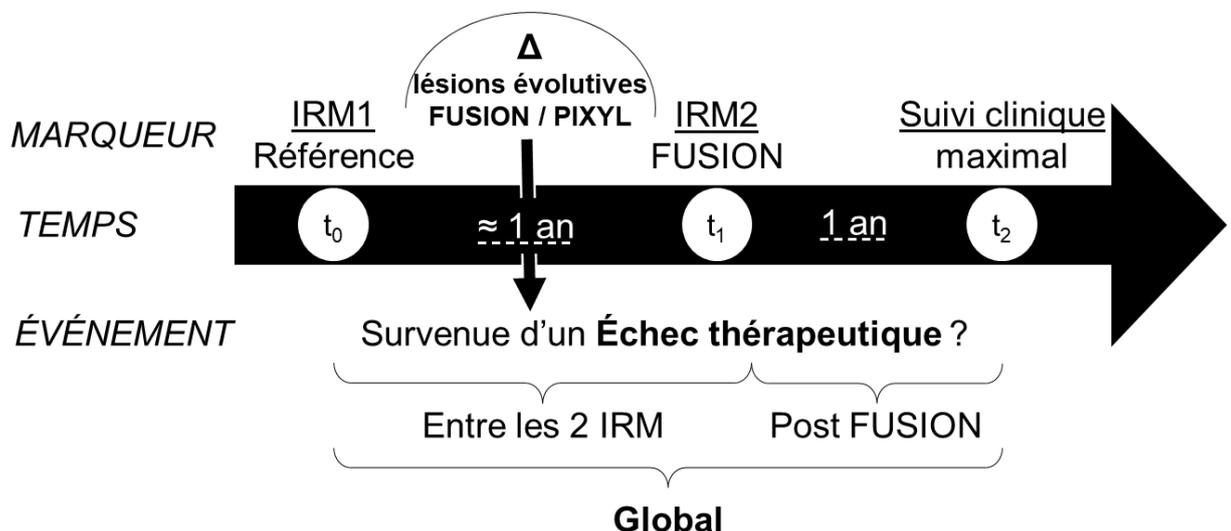


Figure 6. Chronologie des marqueurs de l'étude de comparaison des lésions évolutives selon les méthodes FUSION et PIXYL pour identifier la survenue d'un échec thérapeutique.

### 3. Analyses statistiques

Les indicateurs statistiques descriptifs utilisés étaient, en l'absence de mention contraire, les minima et maxima avec les moyennes  $\pm$  écart-types pour les variables quantitatives continues (âges, durée de temps et volumes) ou les médianes  $\pm$  intervalles interquartiles (iqr) pour les variables quantitatives discrètes (EDSS et nombre de lésions), ou encore le compte avec les proportions exprimées en pourcentages dans le cas de variables qualitatives. Des tests statistiques adaptés aux variables analysées ont été choisis, c'est-à-dire les tests de : Welch ou Wilcoxon-Mann-Whitney selon la normalité des distributions évaluée visuellement et par le test de Shapiro-Wilk pour les variables quantitatives (ce qui n'était le cas que pour l'âge), Fisher pour les variables qualitatives, du log-rank pour les courbes de survie établies par la méthode de Kaplan-Meier et le test de DeLong pour la comparaison des aires sous la courbe ROC sensibilité / spécificité. Les analyses statistiques ont été effectuées avec R 4.0.2 à l'aide du logiciel RStudio 1.3.1073 avec les packages supplémentaires prettyR, ggplot2, survminer, survival, pROC, vcd, psy, boot et MASS (4,12,23,30,35,41,42,44,48,49,51,53). Une valeur de  $p < 0,05$  était retenue comme statistiquement significative et les intervalles de confiance étaient calculés à 95 % (par bootstrap 10000x pour le kappa de Cohen).

## RÉSULTATS

### I. Caractéristiques de la population étudiée

Parmi les 328 patients suivis au CHU de Rouen et inclus dans la cohorte EDMUS, 101 patients avaient eu au moins une IRM cérébrale avec FUSION de séquences FLAIR pour comparaison à la date d'inclusion du 7/02/2019. L'IRM de référence la plus ancienne datait de mai 2015. L'inclusion des patients issus de cette population éligible pour l'étude est illustrée par la Figure 7. Parmi ces patients, 18 n'ont pu être inclus dans l'étude, 8 car l'IRM de référence n'était pas disponible (7 IRM de référence réalisées en externe et 1 non récupérable par problème d'archivage sur le PACS), 2 car suivis pour une neuromyéélite optique (NMO) et 8 autres en raison de problème de qualité des séquences FLAIR (3 avec artefacts et 5 n'ayant pas eu de FLAIR 3D malgré la FUSION). Les 83 patients inclus étaient suivis pour une SEP ( $n = 79$  dont 78 formes RR et 1 PP) ou un syndrome cliniquement isolé (CIS) ( $n = 4$ ).

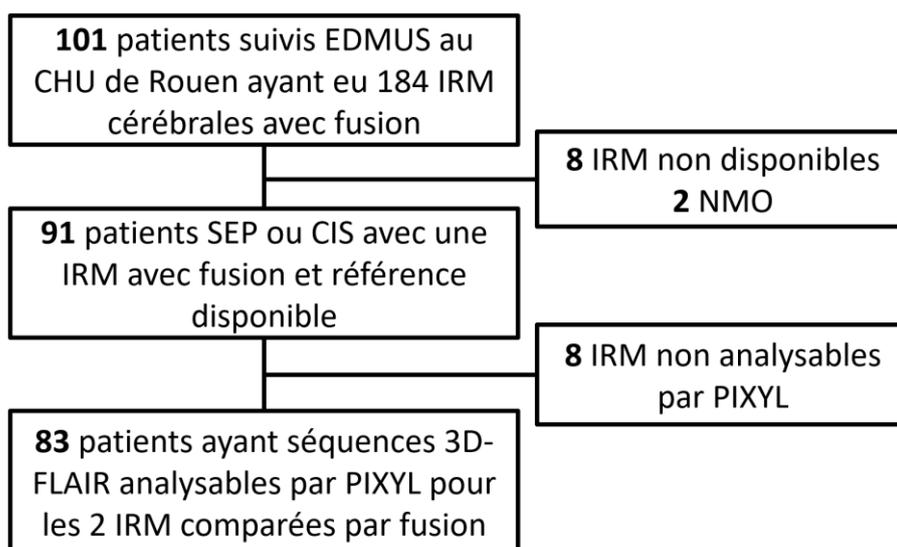


Figure 7. Diagramme de flux des patients inclus dans l'étude.

Le suivi moyen de  $23,8 \pm 5,1$  mois [ $15,0 - 35,7$ ] après l'IRM de référence jusqu'à un an de l'IRM de fusion a permis d'observer un échec thérapeutique chez 33 / 83 patients (39,8 %) survenu  $19,2 \pm 5,9$  mois [ $6,2 - 28,2$ ] après l'IRM de référence ou encore  $8,3 \pm 4,3$  mois [ $0 - 12$ ] après l'IRM de fusion. L'échec thérapeutique était dû à la survenue d'une poussée pour 18 patients (54,5 %), une progression significative de l'EDSS sans poussée pour 4 patients

(12,1 %) et un arrêt du traitement de fond pour inefficacité chez 11 patients (33,3 %). Une majorité des premiers échecs thérapeutiques sont survenus entre les deux IRM ( $n = 25 / 33$  ; 75,8 %) constituant le **sous-groupe de patients avec un échec thérapeutique entre les deux IRM**. Un second échec thérapeutique a été observé après l'IRM de FUSION pour huit des précédents patients soit la moitié des 16 patients constituant le **sous-groupe avec échec thérapeutique après l'IRM de FUSION**. Un seul patient, ayant fait une poussée entre les deux IRM, n'avait pas de suivi clinique ultérieur à l'IRM de FUSION disponible (perdu de vue).

Les caractéristiques générales de la population totale incluse et des groupes avec et sans échec thérapeutique sont données dans le Tableau 4. Aucune différence significative n'existait entre les deux groupes que ce soit en termes d'âge, de sexe, de délai inter-IRM, de durée d'exposition au traitement avant l'IRM ou encore du handicap (EDSS) au moment de l'IRM de référence et de la durée d'évolution de la maladie. L'EDSS différait en revanche au moment de l'IRM de fusion, étant plus élevé dans le groupe avec échec thérapeutique.

Les traitements reçus étaient représentatifs de la diversité des molécules utilisées avec 63 patients traités (75,9 %) lors de l'IRM cérébrale de référence. Il y avait 18 patients (21,7 %) sous traitement immunomodulateur injectable (Interférons et Acétate de glatiramère), 27 patients (32,5 %) sous traitement per os de première ligne (Diméthyl fumarate et Tériflunomide), 18 patients (21,7 %) sous traitement de « haute efficacité » (Alemtuzumab, Natalizumab, Fingolimod, Rituximab et Ocrélizumab). La répartition des traitements ne différait pas significativement entre les deux IRM et est détaillée dans la Figure 8. Pour 21 patients (25,3 %), l'IRM de FUSION correspondait à une rebase à la suite d'un changement de traitement postérieur à l'IRM de référence dont une partie seulement avait présenté un échec thérapeutique ( $n = 12$  ; 57,1 %) entre les deux IRM. Deux de ces patients ont eu un second échec thérapeutique après l'IRM de FUSION. En dehors des échecs thérapeutiques, les changements de traitement entre les deux IRM étaient liés à des introductions de traitement ( $n = 6$ ) et des intolérances ( $n = 3$ , dont un pour sérologie JCV positive). Les échecs thérapeutiques dans un contexte d'IRM de rebase se manifestaient par une durée d'exposition au traitement plus courte dans le sous-groupe de patients avec échec thérapeutique entre les deux IRM (médiane de 13,3 mois  $\pm$  iqr 11 ;  $p = 0,01$ ).

Survenue globale d'un échec thérapeutique depuis l'IRM de référence à un an de l'IRM de FUSION	Population incluse (n = 83)	Groupe avec échec thérapeutique n = 33 (39,8 %)	Groupe sans échec thérapeutique n = 50 (60,2 %)	p*
Âge moyen à l'IRM de référence (années)	39,1 ± 11,9 [16,6-69,5]	39,4 ± 10,9 [20,4-69,5]	38,9 ± 12,5 [16,6-69,3]	0,85
Sexe féminin	63 (75,9 %)	26 (78,8 %)	37 (74,0 %)	0,79
Durée d'évolution de la maladie (années)	6,3 ± 8,2 [0,0-36,7]	6,5 ± 8,0 [0,0-30,5]	6,1 ± 8,4 [0,0-36,7]	0,65
EDSS à l'IRM de référence	1,5 ± 2,0 [0-7]	2,0 ± 3,0 [0-7]	1,3 ± 2,5 [0-7]	0,09
EDSS à l'IRM de fusion	2,0 ± 2,0 [0-7]	2,5 ± 3,5 [0-6]	1,0 ± 1,8 [0-7]	<b>&lt;0,001</b>
Délai entre les 2 IRM (mois)	11,8 ± 5,1 [3,0-23,7]	10,8 ± 5,1 [3,0-23,7]	12,4 ± 5,1 [3,0-23,4]	0,18
Durée médiane ± iqr d'exposition au traitement à l'IRM de fusion (mois)	20,3 ± 28,2 [0-150,2]	14,4 ± 19,8 [0-75,7]	25,6 ± 32,2 [0,3-150,2]	0,055

Tableau 4. Caractéristiques générales dans la population incluse et les groupes avec et sans échec thérapeutique global (\*p, probabilité du test statistique comparant les deux groupes)

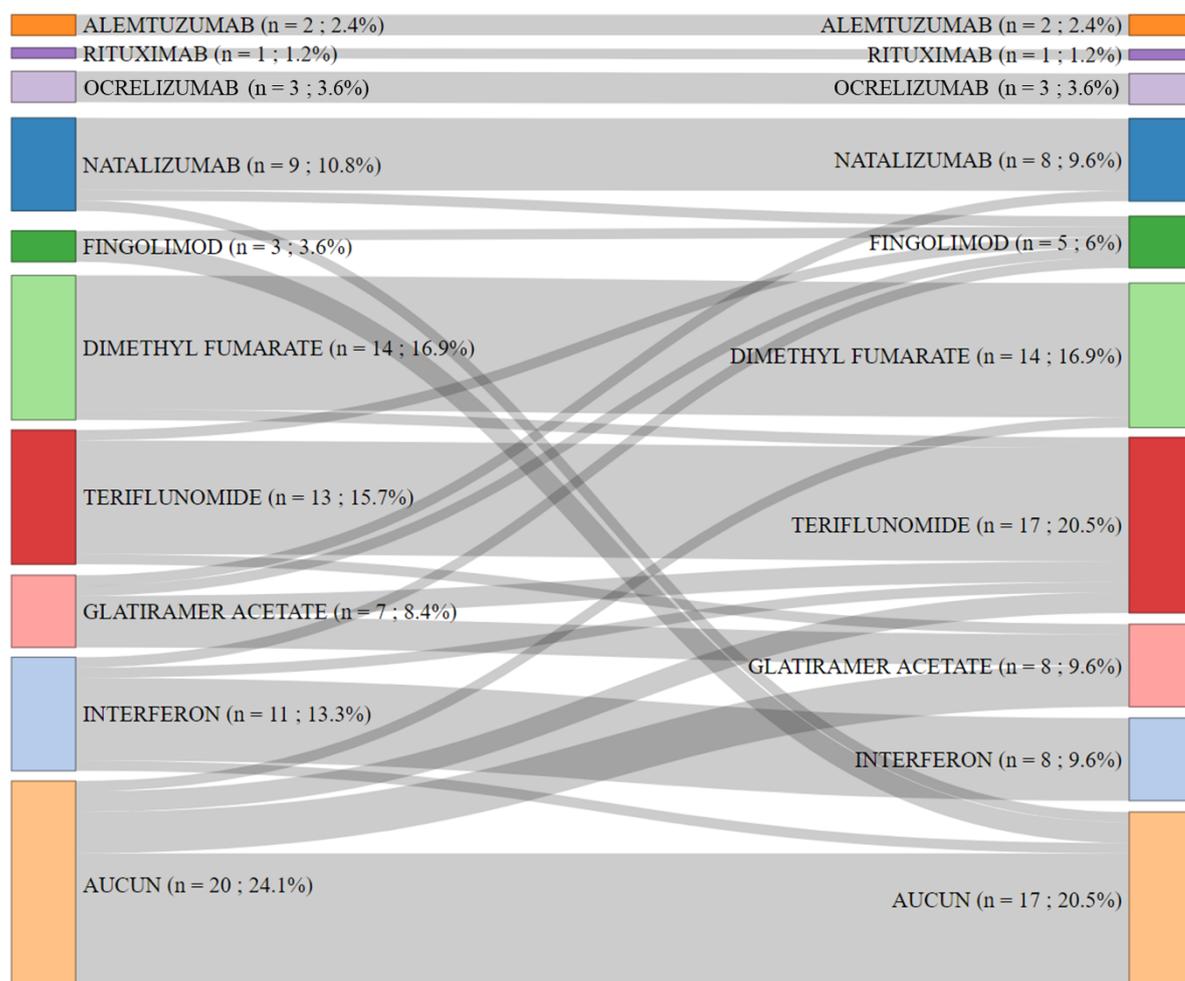


Figure 8. Diagramme de flux de Sankey des traitements pris lors des IRM (à gauche, la répartition des traitements lors de l'IRM de référence et à droite, celle lors de l'IRM de fusion)

## II. Marqueurs radiologiques FUSION et PIXYL

### 1. Par rapport à l'échec thérapeutique global

La charge et le volume lésionnels totaux mesurés par PIXYL lors de l'IRM de FUSION ne différaient pas entre les deux groupes avec une médiane globale de  $40 \pm 36$  lésions [10 - 144] ( $p = 0,99$ ) pour un volume moyen global de  $10,8 \pm 105,7$  ml [0,77 - 58,7] ( $p = 1$ ). Des données radiologiques FUSION et PIXYL ont été synthétisées dans le Tableau 5 pour la population générale et la survenue d'un échec thérapeutique global (entre les deux IRM ou après l'IRM de FUSION). PIXYL apportait des données supplémentaires par rapport à la FUSION avec les dimensions et volumes lésionnels et distinguait les lésions élargies des nouvelles lésions. Une taille seuil pouvait être définie pour filtrer les lésions détectées par PIXYL. Les données présentées dans le tableau portaient donc sur les lésions de toute taille et celles  $\geq 3$  mm de diamètre dans les trois axes.

Il y avait plus d'activité radiologique significative 'FUSION  $\geq 3$  nouvelles lésions' dans le groupe avec échec thérapeutique par rapport au groupe sans échec (30,3 % vs 4,0 % ;  $p = 0,001$ ) avec un nombre total de lésions détectées supérieur (médiane de 1 vs 0 ;  $p = 0,003$ ). En revanche, l'activité radiologique mesurée avec PIXYL ne différait pas statistiquement pour tous les marqueurs ; notamment pour 'la présence d'une ou au moins trois nouvelles lésions' même avec une taille seuil minimale de 3 mm dans les trois axes. En effet, 28 / 33 patients du groupe avec échec (84,9 %) avaient 'au moins une nouvelle lésion PIXYL  $\geq 3$  mm' et 39 / 50 patients dans le groupe sans échec (78 %) ( $p = 0,57$ ). La progression ' $\geq 3$  nouvelles lésions PIXYL  $\geq 3$  mm' concernait 19 / 33 patients du groupe avec échec (57,6 %) par rapport à 19 / 50 patients du groupe sans échec (38 %) ( $p = 0,11$ ). Le nombre de 'lésions nouvelles PIXYL' ne différait faiblement que pour la taille seuil de 3 mm avec une médiane de 3 versus 1 dans les groupes avec versus sans échec ( $p = 0,047$ ). Le nombre de 'lésions élargies' différait sans taille seuil entre les groupes avec et sans échec (médianes respectives de 2 vs 0,5 ;  $p = 0,03$  sans taille seuil et  $p = 0,04$  pour la taille seuil de 3 mm) de même que le nombre de 'lésions évolutives' (médianes de 6 vs 3,5 ;  $p = 0,046$  et médianes de 5 vs 2 ;  $p = 0,028$ ). La variabilité et la dispersion du nombre de nouvelles lésions, reflétée par l'intervalle interquartile (iqr) et les valeurs extrêmes, semblaient plus importante dans le groupe avec échec (valeurs extrêmes du nombre de nouvelles lésions avec taille seuil de 3 mm à 0 – 36 et 0 – 21 respectivement dans les groupes avec et sans échec).

<b>Survenue globale d'un échec thérapeutique depuis l'IRM de référence à un an de l'IRM de FUSION</b>	<b>Population incluse (n = 83)</b>	<b>Groupe avec échec thérapeutique n = 33 (39,8 %)</b>	<b>Groupe sans échec thérapeutique n = 50 (60,2 %)</b>	<b>p*</b>
<b>Activité radiologique FUSION</b>				
Nouvelles lésions	30 (36,1 %)	17 (51,1 %)	13 (26,0 %)	<b>0,02</b>
Progression ≥ 3 nouvelles lésions	12 (14,5 %)	10 (30,3 %)	2 (4,0 %)	<b>0,001</b>
Nombre de nouvelles lésions	0 ± 1 [0-11]	1 ± 3 [0-11]	0 ± 0,8 [0-3]	<b>0,003</b>
<b>Activité radiologique PIXYL</b>				
Nouvelles lésions	73 (88,0 %)	29 (87,9 %)	44 (88,0 %)	1
Lésions élargies	49 (59,0 %)	24 (72,7 %)	25 (50,0 %)	<b>0,044</b>
Lésions évolutives	76 (91,6 %)	31 (93,9 %)	45 (90,0 %)	0,70
Progression ≥ 3 nouvelles lésions	45 (54,2 %)	22 (66,7 %)	23 (46,0 %)	0,08
Nombre de nouvelles lésions	3 ± 6 [0-37]	4 ± 10 [0-37]	2 ± 5 [0-27]	0,10
Nombre de lésions élargies	1 ± 3 [0-18]	2 ± 3 [0-10]	0,5 ± 2 [0-18]	<b>0,03</b>
Nombre total de lésions évolutives	4 ± 8 [0-44]	6 ± 9 [1-44]	3,5 ± 6 [0-41]	<b>0,046</b>
Volume des nouvelles lésions (ml)	0,26 ± 0,56 [0-4,5]	0,42 ± 0,83 [0-4,5]	0,15 ± 0,21 [0-0,8]	0,06
Volume des lésions élargies (ml)	0,96 ± 2,3 [0-13,6]	1,6 ± 3,0 [0-13,6]	0,52 ± 1,6 [0-8,6]	<b>0,005</b>
Volume total de lésions évolutives	1,2 ± 2,8 [0-18,1]	2,1 ± 3,8 [0-18,1]	0,67 ± 1,7 [0-9,2]	<b>0,004</b>
<b>Activité PIXYL seuil ≥ 3 mm 3D</b>				
Nouvelles lésions	67 (80,7 %)	28 (84,9 %)	39 (78,0 %)	0,57
Lésions élargies	48 (57,8 %)	23 (69,7 %)	25 (50,0 %)	0,11
Lésions évolutives	71 (85,5 %)	30 (90,9 %)	41 (82,0 %)	0,35
Progression ≥ 3 nouvelles lésions	38 (45,8 %)	19 (57,6 %)	19 (38,0 %)	0,11
Nombre de nouvelles lésions	2 ± 4 [0-36]	3 ± 7 [0-36]	1 ± 3 [0-21]	<b>0,047</b>
Nombre de lésions élargies	1 ± 3 [0-18]	2 ± 3 [0-10]	0,5 ± 2 [0-18]	<b>0,039</b>
Nombre total de lésions évolutives	3 ± 7 [0-43]	5 ± 7 [0-43]	2 ± 4 [0-35]	<b>0,028</b>
Volume des nouvelles lésions (ml)	0,24 ± 0,55 [0-4,5]	0,4 ± 0,83 [0-4,5]	0,13 ± 0,2 [0-0,8]	<b>0,040</b>
Volume des lésions élargies (ml)	0,96 ± 2,3 [0-13,6]	1,6 ± 3,0 [0-13,6]	0,52 ± 1,6 [0-8,6]	<b>0,005</b>
Volume total de lésions évolutives	1,2 ± 2,8 [0-18,1]	2,0 ± 3,8 [0-18,1]	0,65 ± 1,7 [0-9,1]	<b>0,004</b>

Tableau 5. Activité radiologique FUSION et PIXYL dans la population incluse et les groupes avec et sans échec thérapeutique global (\*p, probabilité du test statistique comparant les deux groupes) (sauf mention contraire, moyenne ± déviation standard pour les variables quantitatives continues ; médiane ± intervalle interquartile pour les variables quantitatives discrètes ; [minimum-maximum] ; nombre et proportion en % pour les variables qualitatives)

De manière similaire au nombre de lésions, les volumes différaient pour les lésions élargies et évolutives sans critère de taille seuil (respectivement  $p = 0,005$  et  $p = 0,004$  pour des volumes moyens respectifs de 0,42 versus 0,15 ml et 2,1 versus 0,67 ml dans les groupes avec et sans échec). Le volume des nouvelles lésions ne différait que pour une taille seuil de 3 mm (0,4 vs 0,13 ml ;  $p = 0,04$ ). Pour les volumes, la dispersion semblait plus importante pour les lésions élargies (0 à 13,6 ml) par rapport aux nouvelles lésions (0 à 4,5 ml si taille seuil de 3 mm). Il s'agissait du volume total des lésions détectées comme élargies et pas du volume d'élargissement qui n'était pas recueilli dans l'étude. Le volume moyen des lésions élargies

(0,96 ml) n'était cependant pas significativement plus important que celui des nouvelles lésions (0,24 ml) ( $p = 0,87$ ). Les comptes et volumes des lésions élargies étaient identiques sans et avec taille seuil de 3 mm à l'exception d'une lésion chez un patient. Ceci signifiait qu'une lésion élargie PIXYL avait presque toujours au moins une taille minimale de 3 mm dans les trois axes et un volume de 9 mm<sup>3</sup>.

PIXYL retrouvait nettement plus de nouvelles lésions que FUSION (médiane de 3 vs 0 ;  $p < 0,001$ ) même pour une taille seuil  $\geq 3$  mm dans les trois axes (médiane de 2 vs 0 ;  $p < 0,001$ ). Une faible corrélation était constatée entre le nombre de nouvelles lésions obtenu par les deux méthodes (coefficient de corrélation Tau de Kendall de 0,24 et 0,26 sans et avec une taille seuil de 3 mm respectivement ;  $p = 0,007$  et  $p = 0,004$ ). Il en résultait une activité radiologique PIXYL plus importante avec présence d'au moins une lésion évolutive (souvent nouvelle) chez une majorité de patients ( $n = 76 / 83$  ; 91,6 %) quel que soit le groupe, même si cela tendait à être moins marqué dans le groupe sans échec thérapeutique dans le cas d'une taille seuil de 3 mm (41 / 50 ; 82 % vs 30 / 33 ; 90,9 % dans le groupe avec échec,  $p = 0,35$ ). Les seuils  $n \geq 12$  nouvelles lésions ou  $n \geq 17$  lésions évolutives PIXYL permettaient d'obtenir la prévalence la plus proche de celle du seuil classique  $n \geq 3$  nouvelles lésions FUSION chez 12 / 83 patients (14,5 %). C'est-à-dire que la prévalence était de 11 / 83 patients (13,3 %) dont 8 dans le groupe avec échec thérapeutique et 12 / 83 (14,5 %) dont 8 dans le groupe avec échec pour les seuils respectifs de lésions nouvelles et évolutives. Ces seuils étaient abaissés respectivement à  $n \geq 9$  et  $n \geq 15$  lorsque l'on retenait la taille seuil des lésions nouvelles et évolutives  $\geq 3$  mm dans les 3 axes. C'est-à-dire que la prévalence était de 13 et 12 / 83 patients (15,7 % et 14,5 % respectivement) dont 8 et 8 avec échec. Néanmoins, la répartition des lésions suivant ces seuils d'équivalence ne différait entre les deux groupes selon l'échec que pour le seuil  $n \geq 12$  nouvelles lésions sans taille seuil ( $p = 0,02$ ). Les décomptes de lésions PIXYL élargies et évolutives n'étaient pas plus fortement corrélés avec le nombre de nouvelles lésions FUSION (Tau de Kendall de 0,32 et 0,28 respectivement avec taille seuil de 3 mm ;  $p < 0,001$  et  $p = 0,001$ ).

Le nombre de nouvelles lésions identifiées par PIXYL par territoire cérébral ne différait selon l'échec thérapeutique global que pour les lésions infratentorielles sans critère de taille seuil comme pour le seuil de 3 mm (médiane de  $0 \pm 1$  [0 - 11] dans le groupe avec échec versus  $0 \pm 0$  [0 - 1] dans le groupe sans échec ;  $p = 0,02$ ). Le nombre médian de lésions par localisation

était faible de 0 ou 1 avec des maximas à 21 pour les lésions juxta-corticales (20 si taille seuil de 3 mm), 11 en infratentorial (11), 8 en périventriculaire (7) et 7 dans la substance blanche profonde (6).

Le nombre de nouvelles lésions FUSION différait selon la présence d'un traitement au moment de l'IRM de référence. En effet, les patients sans traitement avaient une médiane de  $1 \pm 2,25$  nouvelles lésions par rapport à  $0 \pm 1$  pour les patients sous traitement ( $p = 0,04$ ). La proportion de patients ayant au moins une ou trois nouvelles lésions FUSION ne différait pas selon la prise d'un traitement à l'IRM de référence ( $p > 0,06$ ). Aucun marqueur radiologique PIXYL ni proportion de patients selon les seuils utilisés dans le Tableau 5 ne différait selon le traitement reçu lors de l'IRM de référence ( $p$  tous supérieurs à 0,45).

## 2. Par rapport à l'échec thérapeutique entre les deux IRM

Dans le sous-groupe de patients ayant eu un échec thérapeutique entre les deux IRM, les activités radiologiques FUSION et PIXYL étaient similaires à celles dans le groupe avec échec thérapeutique global. Les différences entre les groupes avec et sans échec entre les deux IRM étaient juste un peu plus significatives pour l'activité PIXYL que selon l'échec global (Tableau 6).

<b>Survenue d'un échec thérapeutique entre les deux IRM</b>	<b>Échec thérapeutique entre les 2 IRM</b> <i>n</i> = 25 (30,1 %)	<b>Pas d'échec thérapeutique entre les 2 IRM</b> <i>n</i> = 58 (69,9 %)	<i>p</i> *
<b>Activité radiologique FUSION</b>			
Nouvelles lésions	15 (60 %)	15 (25,9 %)	<b>&lt;0,001</b>
Progression ≥ 3 nouvelles lésions	10 (40 %)	2 (3,4 %)	<b>0,006</b>
Nombre de nouvelles lésions	1 ± 3 [0-11]	0 ± 0,75 [0-3]	<b>&lt;0,001</b>
<b>Activité radiologique PIXYL</b>			
Nouvelles lésions	23 (92 %)	50 (86,2 %)	0,716
Lésions élargies	21 (84 %)	28 (48,3 %)	<b>0,003</b>
Lésions évolutives	24 (96 %)	52 (89,7 %)	0,669
Progression ≥ 3 nouvelles lésions	15 (60 %)	25 (43,1 %)	0,231
Nombre de nouvelles lésions	4 ± 13 [0-37]	2 ± 5 [0-27]	<b>0,031</b>
Nombre de lésions élargies	3 ± 3 [0-10]	0 ± 2 [0-18]	<b>0,001</b>
Nombre total de lésions évolutives	7 ± 14 [0-44]	4 ± 6 [0-41]	<b>0,009</b>
Volume des nouvelles lésions (ml)	0,52 ± 0,94 [0-4,54]	0,15 ± 0,2 [0-0,81]	<b>0,015</b>
Volume des lésions élargies (ml)	2,1 ± 3,36 [0-13,56]	0,47 ± 1,46 [0-8,55]	<b>&lt;0,001</b>
Volume total de lésions évolutives	2,62 ± 4,19 [0-18,1]	0,62 ± 1,56 [0-9,16]	<b>&lt;0,001</b>
<b>Activité PIXYL seuil ≥ 3 mm 3D</b>			
Nouvelles lésions	23 (92 %)	44 (75,9 %)	0,13
Lésions élargies	20 (80 %)	28 (48,3 %)	<b>0,008</b>
Lésions évolutives	24 (96 %)	47 (81 %)	0,096
Progression ≥ 3 nouvelles lésions	13 (52 %)	16 (27,6 %)	<b>0,045</b>
Nombre de nouvelles lésions	4 ± 11 [0-36]	1 ± 3 [0-21]	<b>0,006</b>
Nombre de lésions élargies	3 ± 3 [0-10]	0 ± 2 [0-18]	<b>0,003</b>
Nombre total de lésions évolutives	6 ± 13 [0-43]	2 ± 4 [0-35]	<b>0,002</b>
Volume des nouvelles lésions (ml)	0,49 ± 0,93 [0-4,53]	0,13 ± 0,19 [0-0,8]	<b>0,007</b>
Volume des lésions élargies (ml)	2,1 ± 3,36 [0-13,56]	0,47 ± 1,46 [0-8,55]	<b>&lt;0,001</b>
Volume total de lésions évolutives	2,59 ± 4,18 [0-18,09]	0,6 ± 1,55 [0-9,09]	<b>&lt;0,001</b>

Tableau 6. Activité radiologique FUSION et PIXYL dans les groupes avec et sans échec thérapeutique survenu entre les deux IRM (\**p*, probabilité du test statistique comparant les deux groupes) (moyenne ± déviation standard pour les variables quantitatives continues ; médiane ± intervalle interquartile pour les variables quantitatives discrètes ; [minimum-maximum] ; nombre et proportion en % pour les variables qualitatives)

### 3. Par rapport à l'échec thérapeutique après l'IRM de FUSION

Dans le sous-groupe de patients ayant eu un échec thérapeutique dans l'année suivant l'IRM de FUSION, aucune différence significative de l'activité radiologique mesurée par FUSION et PIXYL ne ressortait (Tableau 7).

<b>Survenue d'un échec thérapeutique à 1 an de l'IRM de FUSION</b>	<b>Échec thérapeutique après la FUSION</b> <i>n</i> = 16 (19,3 %)	<b>Pas d'échec thérapeutique après la FUSION</b> <i>n</i> = 67 (80,7 %)	<i>p</i> *
<b>Activité radiologique FUSION</b>			
Nouvelles lésions	8 (50 %)	22 (32,8 %)	0,2
Progression ≥ 3 nouvelles lésions	4 (25 %)	8 (11,9 %)	0,25
Nombre de nouvelles lésions	1 ± 2,25 [0-11]	0 ± 1 [0-4]	0,074
<b>Activité radiologique PIXYL</b>			
Nouvelles lésions	14 (87,5 %)	59 (88,1 %)	1
Lésions élargies	10 (62,5 %)	39 (58,2 %)	1
Lésions évolutives	15 (93,8 %)	61 (91 %)	1
Progression ≥ 3 nouvelles lésions	11 (68,8 %)	29 (43,3 %)	0,095
Nombre de nouvelles lésions	5 ± 3,75 [0-28]	2 ± 6 [0-37]	0,254
Nombre de lésions élargies	1,5 ± 3,25 [0-7]	1 ± 3 [0-18]	0,52
Nombre total de lésions évolutives	5,5 ± 6 [0-35]	4 ± 8 [0-44]	0,299
Volume des nouvelles lésions (ml)	0,29 ± 0,38 [0-1,36]	0,25 ± 0,6 [0-4,54]	0,318
Volume des lésions élargies (ml)	0,93 ± 1,24 [0-3,67]	0,97 ± 2,51 [0-13,56]	0,318
Volume total de lésions évolutives	1,22 ± 1,49 [0-5,03]	1,22 ± 3,01 [0-18,1]	0,157
<b>Activité PIXYL seuil ≥ 3 mm 3D</b>			
Nouvelles lésions	13 (81,3 %)	54 (80,6 %)	1
Lésions élargies	10 (62,5 %)	38 (56,7 %)	0,782
Lésions évolutives	14 (87,5 %)	57 (85,1 %)	1
Progression ≥ 3 nouvelles lésions	7 (43,8 %)	22 (32,8 %)	0,56
Nombre de nouvelles lésions	3 ± 4 [0-21]	2 ± 3 [0-36]	0,359
Nombre de lésions élargies	1,5 ± 3,25 [0-7]	1 ± 3 [0-18]	0,488
Nombre total de lésions évolutives	4 ± 6,5 [0-28]	3 ± 5,5 [0-43]	0,437
Volume des nouvelles lésions (ml)	0,26 ± 0,37 [0-1,32]	0,24 ± 0,59 [0-4,53]	0,408
Volume des lésions élargies (ml)	0,93 ± 1,24 [0-3,67]	0,97 ± 2,51 [0-13,56]	0,294
Volume total de lésions évolutives	1,19 ± 1,48 [0-4,98]	1,2 ± 3 [0-18,09]	0,199

Tableau 7. Activité radiologique FUSION et PIXYL dans les groupes avec et sans échec thérapeutique survenu après l'IRM de FUSION (\**p*, valeur de la probabilité du test statistique comparant les deux groupes) (moyenne ± déviation standard pour les variables quantitatives continues ; médiane ± intervalle interquartile pour les variables quantitatives discrètes ; [minimum-maximum] ; nombre et proportion en % pour les variables qualitatives)

Concernant l'échec thérapeutique global ou entre les deux IRM :

Les marqueurs radiologiques FUSION (présence d'une nouvelle lésion, nombre de nouvelle lésion et progression  $\geq 3$  nouvelles lésions) étaient donc tous supérieurs dans le groupe avec échec thérapeutique. Les marqueurs radiologiques PIXYL différaient selon l'échec global pour une taille seuil de 3 mm mais les seuils validés pour la FUSION (présence d'une lésion ou au moins trois lésions, qu'elles soient nouvelles, élargies ou évolutives) ne permettaient pas de distinguer le groupe avec ou sans échec global. La présence de lésions PIXYL élargies ou d'au moins trois nouvelles lésions  $\geq 3$  mm différaient selon l'échec entre les deux IRM. Les décomptes de lésions FUSION et PIXYL n'étaient que peu corrélés (Tau de Kendall  $< 0,5$ ).

Concernant l'échec thérapeutique après l'IRM de FUSION :

Aucun des marqueurs radiologiques FUSION et PIXYL ne différaient entre les groupes avec et sans échec thérapeutique.

### **III. Comparaison des performances de la mesure d'activité radiologique FUSION et PIXYL pour identifier l'échec thérapeutique**

#### **1. Par rapport à l'échec thérapeutique global**

Les performances des marqueurs d'activité radiologique PIXYL et FUSION pour identifier l'échec thérapeutique ont été explorées. Ces données ont été synthétisées dans le Tableau 8 et le Tableau 9.

Le nombre d'au moins trois nouvelles lésions détectées par FUSION a été pris comme référence avec une excellente spécificité de 96 % (2 faux positifs, l'un dans le cadre d'une rebaseline et l'autre d'une activité radiologique infraclinique tolérée en raison de la stabilité clinique et de facteurs personnels du patient concernant le choix du traitement) mais une sensibilité de 30,3 % résultant en une exactitude pour prédire l'échec thérapeutique de 69,9 %. Le même seuil appliqué aux comptes de nouvelles lésions PIXYL avec et sans filtre de taille  $\geq 3$  dans les trois axes (3D) obtient une sensibilité meilleure à respectivement 66,7 % et 57,6 % ( $p = 0,006$  et  $p = 0,046$ ) mais la spécificité est nettement moindre à 54 % et 62 % ( $p < 0,001$ ). L'exactitude calculée à 59,0 % et 60,2 % ne différait pas significativement de celle  $\geq 3$  nouvelles lésions FUSION ( $p = 0,2$  et  $p = 0,3$ ).

Des marqueurs composites ont été construits pour essayer de combiner le maximum d'informations fournies par PIXYL. Le degré de certitude de la segmentation des lésions nouvelles ou élargies a permis d'obtenir un '**nombre de lésions pondéré**' calculé comme le produit du degré de certitude (de 0 à 1) par le nombre de lésions d'un type défini. Un '**ratio des différents volumes**' était calculé comme le rapport du volume divisé par le nombre de lésions considérées (nouvelles, élargies ou évolutives).

**La comparaison des aires sous la courbe ROC sensibilité / spécificité (AUC) ne permettait pas de mettre en évidence de différence statistique pour identifier l'échec thérapeutique entre le compte de nouvelles lésions selon la méthode de référence FUSION et les différents indicateurs radiologiques PIXYL étudiés (Tableau 8).**

Le nombre de nouvelles lésions FUSION permettait d'identifier un échec thérapeutique global avec une AUC de 0,668 comparé au nombre de nouvelle lésions PIXYL ayant une AUC de 0,605 ( $p = 0,39$ ) sans taille seuil et 0,628 ( $p = 0,57$ ) si taille seuil  $\geq 3$  mm.

Les AUC de certains marqueurs pour les nouvelles lésions PIXYL (nombre sans taille seuil, ce nombre pondéré par le degré de certitude, volume et ratio des volumes, même pour une taille seuil de 3 mm pour ce dernier) pouvaient être considérées comme mauvaises, incluant 0,5 dans l'intervalle de confiance à 95 % (valeurs grisées dans le Tableau 8). Les indicateurs PIXYL comportant les volumes des lésions élargies tendaient à avoir les AUC les plus hautes. La meilleure AUC était celle pour le 'ratio du volume évolutif / nombre de lésions évolutives' calculée à 0,70. Elle ne différait pas de l'AUC selon le nombre de lésions FUSION ( $p = 0,6$ ).

Ces résultats apportaient d'autres informations intéressantes. Les meilleurs seuils (valeur  $\geq n$ ) des courbes sensibilité / spécificité, définis comme ayant la meilleure somme sensibilité / spécificité, ont été retenus pour chaque marqueur radiologique PIXYL dans le Tableau 8 et le Tableau 9.

Le seuil  $\geq 2$  nouvelles lésions détectées par FUSION semblait avoir un meilleur rapport sensibilité / spécificité (0,424 / 0,92) et une plus grande exactitude (0,723) que le seuil classique  $\geq 3$  (sensibilité / spécificité à 0,303 / 0,96) (exactitude de 0,699) bien que ne différant pas significativement ( $p > 0,5$  pour tous les indicateurs statistiques). Tous les seuils retenus pour les variables PIXYL étaient associés à des répartitions significativement différentes entre les groupes avec et sans échec thérapeutique ( $p < 0,045$ ) hormis le nombre de nouvelles lésions sans taille seuil  $\geq 3$  ( $p = 0,075$ ) (Tableau 9).

Dans les tableaux suivants, les valeurs supérieures ou égales à celles de la référence selon le nombre de lésions FUSION (seuil de 3) sont en gras et soulignées si statistiquement différentes (test de Fisher).

Marqueur radiologique	AUC [IC95%]	<i>p</i>	Seuil	Se	Spé	VPP	VPN	Exact
<b>Nouvelle lésion FUSION</b>	<b>0,668</b> <b>[0,558-0,779]</b>	1	3	<b>0,303</b>	<b>0,96</b>	<b>0,833</b>	<b>0,676</b>	<b>0,699</b>
Nouvelle lésion FUSION	<b>0,668</b> <b>[0,558-0,779]</b>	1	2	<b>0,424</b>	0,92	0,778	<b>0,708</b>	<b>0,723</b>
Nouvelle lésion	0,605 [0,478-0,733]	0,392	3	<b>0,667</b>	<u>0,54</u>	<u>0,489</u>	<b>0,711</b>	0,59
Nouvelle lésion 3D*	0,628 [0,503-0,753]	0,565	5	<b>0,394</b>	0,82	0,591	0,672	0,651
Lésion élargie	0,639 [0,52-0,758]	0,678	3	<b>0,485</b>	<u>0,78</u>	0,593	<b>0,696</b>	0,663
Lésion élargie 3D*	0,629 [0,509-0,749]	0,565	3	<b>0,485</b>	<u>0,78</u>	0,593	<b>0,696</b>	0,663
Lésion évolutive	0,63 [0,505-0,754]	0,586	5	<b>0,636</b>	<u>0,6</u>	0,512	<b>0,714</b>	0,614
Lésion évolutive 3D*	0,643 [0,519-0,766]	0,709	5	<b>0,545</b>	<u>0,7</u>	0,545	<b>0,7</b>	0,639
Nouvelle lésion pondérée	0,617 [0,491-0,743]	0,481	0,77	<b>0,818</b>	<u>0,44</u>	0,491	<b>0,786</b>	0,59
Lésion élargie pondérée	0,648 [0,528-0,768]	0,769	0,77	<b>0,636</b>	<u>0,68</u>	0,568	<b>0,739</b>	0,663
Volume nouvelles lésions	0,625 [0,498-0,752]	0,51	106	<b>0,606</b>	<u>0,64</u>	0,526	<b>0,711</b>	0,627
Volume nouvelles lésions 3D*	0,633 [0,507-0,76]	0,59	67	<b>0,667</b>	<u>0,62</u>	0,537	<b>0,738</b>	0,639
Volume lésions élargies	<b>0,676</b> <b>[0,557-0,795]</b>	0,906	181	<b>0,606</b>	<u>0,76</u>	0,625	<b>0,745</b>	0,699
Volume lésions élargies 3D*	0,667 [0,546-0,787]	0,982	181	<b>0,606</b>	<u>0,76</u>	0,625	<b>0,745</b>	0,699
Volume évolutif	<b>0,688</b> <b>[0,568-0,808]</b>	0,75	303	<b>0,667</b>	<u>0,7</u>	0,595	<b>0,761</b>	0,687
Volume évolutif 3D*	<b>0,687</b> <b>[0,567-0,807]</b>	0,761	293	<b>0,667</b>	<u>0,7</u>	0,595	<b>0,761</b>	0,687
Ratio volume / nouvelle lésion	0,608 [0,48-0,737]	0,319	29,9	<b>0,515</b>	<u>0,74</u>	0,567	<b>0,698</b>	0,651
Ratio volume / nouvelle lésion 3D*	0,615 [0,489-0,741]	0,381	35,3	<b>0,515</b>	<u>0,74</u>	0,567	<b>0,698</b>	0,651
Ratio volume / lésion élargie	<b>0,679</b> <b>[0,561-0,797]</b>	0,871	85,8	<b>0,576</b>	<u>0,74</u>	0,594	<b>0,725</b>	0,675
Ratio volume / lésion élargie 3D*	<b>0,67</b> <b>[0,55-0,79]</b>	0,982	85,8	<b>0,576</b>	<u>0,74</u>	0,594	<b>0,725</b>	0,675
Ratio volume évolutif / lésion évolutive	<b>0,7</b> <b>[0,581-0,819]</b>	0,596	35	<b>0,788</b>	<u>0,58</u>	0,553	<b>0,806</b>	0,663
Ratio volume évolutif / lésion évolutive 3D*	<b>0,697</b> <b>[0,579-0,815]</b>	0,632	48	<b>0,727</b>	<u>0,64</u>	0,571	<b>0,78</b>	0,675

Tableau 8. Performances de différents marqueurs radiologiques PIXYL comparées à celles du nombre de lésion comptées en FUSION pour identifier la survenue d'un échec thérapeutique (\*3D = taille seuil  $\geq 3$  mm dans les trois axes ; Volumes en mm<sup>3</sup> ; AUC : aire sous la courbe ROC et intervalle de confiance à 95 % ; *p* : probabilité d'une différence entre l'AUC du

marqueur radiologique étudié et l'AUC du nombre de nouvelles lésions FUSION (méthode DeLong) ; Seuil optimal  $\geq n$  (meilleure somme Se + Spé) ; Se : sensibilité ; Spé : spécificité ; VPP : valeur prédictive positive ; VPN : valeur prédictive négative ; Exactitude de la classification). Les AUC dont l'intervalle de confiance comprenait 0,5 sont sur fond gris.

Marqueur radiologique	<i>n</i>	<i>p</i>	OR [IC95%]	VP	VN	FP	FN
Nouvelle lésion FUSION	2	<b>&lt;0,001</b>	8,2 [2,2-38,8]	14	46	4	19
Nouvelle lésion FUSION	3	<b>0,001</b>	10,1 [1,9-102,6]	10	48	2	23
Nouvelle lésion	3	0,075	2,3 [0,9-6,5]	22	27	23	11
Nouvelle lésion 3D*	5	<b>0,042</b>	2,9 [1-9,2]	13	41	9	20
Lésion élargie	3	<b>0,017</b>	3,3 [1,2-9,7]	16	39	11	17
Lésion élargie 3D*	3	<b>0,017</b>	3,3 [1,2-9,7]	16	39	11	17
Lésion évolutive	5	<b>0,045</b>	2,6 [1-7,2]	21	30	20	12
Lésion évolutive 3D*	5	<b>0,039</b>	2,8 [1-7,7]	18	35	15	15
Nouvelle lésion pondérée	0,77	<b>0,018</b>	3,5 [1,1-12,2]	27	22	28	6
Lésion élargie pondérée	0,77	<b>0,007</b>	3,7 [1,3-10,4]	21	34	16	12
Volume nouvelles lésions	106	<b>0,042</b>	2,7 [1-7,5]	20	32	18	13
Volume nouvelles lésions 3D*	67	<b>0,014</b>	3,2 [1,2-9,2]	22	31	19	11
Volume lésions élargies	181	<b>0,001</b>	4,8 [1,7-14,1]	20	38	12	13
Volume lésions élargies 3D*	181	<b>0,001</b>	4,8 [1,7-14,1]	20	38	12	13
Volume évolutif	303	<b>0,001</b>	4,6 [1,7-13,4]	22	35	15	11
Volume évolutif 3D*	293	<b>0,001</b>	4,6 [1,7-13,4]	22	35	15	11
Ratio volume / nouvelle lésion	29,9	<b>0,021</b>	3 [1,1-8,5]	17	37	13	16
Ratio volume / nouvelle lésion 3D*	35,3	<b>0,021</b>	3 [1,1-8,5]	17	37	13	16
Ratio volume / lésion élargie	85,8	<b>0,006</b>	3,8 [1,4-10,9]	19	37	13	14
Ratio volume / lésion élargie 3D*	85,8	<b>0,006</b>	3,8 [1,4-10,9]	19	37	13	14
Somme ratio volumes / lésion nouvelle et élargie	215,4	<b>0,004</b>	4,2 [1,5-12,7]	17	40	10	16
Ratio volume évolutif / lésion évolutive	35	<b>0,001</b>	5 [1,7-16,4]	26	29	21	7
Ratio volume évolutif / lésion évolutive 3D*	48	<b>0,002</b>	4,6 [1,7-14,0]	24	32	18	9

Tableau 9. Application des seuils de performance optimale  $\geq n$  aux marqueurs radiologiques PIXYL et comparaison de la différence selon l'échec thérapeutique (\*3D = taille seuil  $\geq 3$  mm dans les trois axes ; Volumes en mm<sup>3</sup> ; *p* et odds ratio OR avec intervalle de confiance à 95 % selon le test de Fisher ; VP : vrais positifs ; VN : vrais négatifs ; FP : faux positifs ; FN : faux négatifs)

La plupart des indicateurs PIXYL avaient une sensibilité plus élevée que le seuil '  $\geq 3$  nouvelles lésions FUSION' ( $p < 0,04$  au minimum pour les seuils indiqués dans le Tableau 8 et soulignés). Ces sensibilités différaient significativement de celle de la référence FUSION

(30,3 %) s'étendaient de 57,6 % pour les 'ratios des volumes / lésion élargies' à 81,8 % pour le 'nombre de nouvelles lésions pondérées'.

La spécificité était moindre avec PIXYL par rapport au seuil de référence FUSION (96 %) sauf pour le seuil '≥ 3 nouvelles lésions de taille seuil de 3 mm' où elle ne différait pas (82 % ;  $p = 0,051$ ). Pour les seuils optimaux des autres marqueurs PIXYL, elle variait de 44 % pour le 'nombre de nouvelles lésions pondérées' ≥ 0,77 à 78 % pour le 'nombre de lésions élargies' ≥ 3. Les valeurs prédictives positives PIXYL, bien que semblant plus faibles, ne différaient pas par rapport à FUSION (83,3 %) sauf pour le seuil ≥ 3 'nouvelles lésions sans taille seuil' où elle était limite plus faible (48,9 % ;  $p = 0,049$ ).

Les VPP des seuils des autres marqueurs variaient de 49,1 % pour le 'nombre de nouvelles lésions pondérées' ≥ 0,77 à 62,5 % pour le 'volume des lésions élargies' ≥ 181 mm<sup>3</sup> sans et avec la taille seuil de 3 mm. Les valeurs prédictives négatives tendaient à être plus élevées avec PIXYL par rapport au seuil de référence de FUSION (67,6 %), reflet de la meilleure sensibilité, mais sans significativité ( $p > 0,18$  au minimum). Les VPN PIXYL s'étendaient de 67,2 % pour le 'nombre de nouvelles lésions PIXYL de taille seuil 3 mm' ≥ 5 à 80,6 % pour le 'ratio du volume des lésions évolutives' ≥ 35.

La meilleure exactitude était de 72,3 % au seuil ≥ 2 'nouvelles lésions FUSION'. Aucune différence n'était significative pour les exactitudes ( $p > 0,49$  au minimum). Celles de PIXYL variaient de 59 % pour le 'nombre de nouvelles lésions' ≥ 3 (exactitude identique pour le seuil ≥ 0,77 lésion pondérée) à 69,9 % pour le 'volume des lésions élargies' ≥ 181 mm<sup>3</sup>.

Les courbes ROC sensibilité/spécificité des principaux marqueurs radiologiques illustrant les résultats présentés précédemment sont disponibles dans les Annexes ([Annexe 2](#) à [Annexe 7](#)).

Le nombre de nouvelles lésions PIXYL de taille seuil ≥ 3 mm obtenait des performances semblables au nombre de lésions FUSION avec un seuil optimal ≥ 5 pour PIXYL par rapport au seuil ≥ 3 utilisé en routine pour FUSION. La meilleure exactitude était obtenue avec le seuil ≥ 2 lésions FUSION (72,3 %) même si elle ne différait pas statistiquement des exactitudes des autres seuils optimaux FUSION et PIXYL.

## 2. Par rapport à l'échec thérapeutique entre les deux IRM

Dans ce sous-groupe, les performances d'identification de l'échec thérapeutique survenu entre les deux IRM ressemblent à celles rapportées ci-dessus dans le groupe global et sont disponibles dans l'Annexe 8.

L'AUC du nombre de nouvelles lésions FUSION était de 0,717 [0,596 - 0,838]. Le seuil  $\geq 3$  lésions FUSION était très spécifique (96,6 %) avec une sensibilité de 40 %. Son exactitude de 79,5 % était la meilleure de tous les marqueurs radiologiques étudiés. Le seuil  $\geq 2$  lésions FUSION obtenait des performances proches avec une sensibilité de 48 %, une spécificité de 89,7 % et une exactitude de 74,7 %. Le seuil optimal  $\geq 5$  nouvelles lésions PIXYL de taille seuil  $\geq 3$  mm avait une sensibilité de 48 %, une spécificité de 82,8 % et une exactitude de 72,3 %. Ces indicateurs ne différaient pas statistiquement sauf pour la spécificité moindre pour PIXYL ( $p = 0,03$ ). Les AUC des différents marqueurs radiologiques PIXYL n'étaient pas significativement différentes de celle de FUSION et variaient de 0,643 [0,515 - 0,771] pour le ratio volume / nouvelle lésion 3D à au mieux 0,771 [0,657 - 0,885] pour le ratio volume évolutif / lésion évolutive. Les AUC du nombre de nouvelles lésions PIXYL sans et avec taille seuil de 3 mm étaient de 0,649 [0,513 - 0,786] et 0,689 [0,561 - 0,817]. Elles semblaient donc légèrement inférieures à celle de FUSION (0,717).

Certains marqueurs PIXYL avaient plus de sensibilité que la FUSION (40 %) comme le nombre ( $\geq 1$ ) et le volume ( $\geq 181 \text{ mm}^3$ ) des lésions élargies respectivement à 84 % et 72 % ( $p = 0,003$  et  $p = 0,045$ ). Néanmoins les spécificités des meilleurs seuils PIXYL étaient toutes inférieures à FUSION  $\geq 3$  nouvelles lésions, sauf pour les nombres de nouvelles lésions PIXYL sans taille seuil  $\geq 11$  et pondérées  $\geq 6$ , où elles étaient similaires, respectivement à 91,4 % et 94,8 % versus 96,6 % pour FUSION ( $p = 0,44$  et  $p = 1$ ). La sensibilité de ces seuils PIXYL était également proche à 36 % versus 40 % pour FUSION ( $p = 1$ ). Il s'agissait des deux seuils optimaux PIXYL ayant tous leurs indicateurs statistiques significativement non différents de FUSION. Les exactitudes étaient de 61,4 % et 74,7 % pour les seuils optimaux PIXYL ayant une bonne sensibilité, c'est-à-dire du nombre ( $\geq 1$ ) et du volume ( $\geq 181 \text{ mm}^3$ ) des lésions élargies. Elle était significativement inférieure à celle du seuil  $\geq 3$  nouvelles lésions FUSION de 79,5 % pour le seuil  $\geq 1$  lésion élargie PIXYL ( $p = 0,02$ ) et similaire pour le volume  $\geq 181 \text{ mm}^3$  ( $p = 0,58$ ).

Les marqueurs PIXYL n'avaient donc pas de performance globale meilleure que la référence FUSION  $\geq 3$  nouvelles lésions. Les seuils PIXYL  $\geq 11$  nouvelles lésions et  $\geq 6$  nouvelles lésions pondérées avaient des performances similaires à FUSION pour tous les indicateurs statistiques. Le seuil optimal  $\geq 5$  nouvelles lésions PIXYL de taille seuil  $\geq 3$  mm avait une spécificité moindre que FUSION comme la majorité des seuils PIXYL. Le meilleur moyen d'augmenter légèrement la sensibilité sans perdre trop de spécificité était de fixer un seuil  $\geq 2$  nouvelles lésions pour FUSION (pas de différence significative néanmoins, sensibilité 48 % et spécificité 89,7 %).

### 3. Par rapport à l'échec thérapeutique après l'IRM de FUSION

Dans ce sous-groupe, les performances d'identification et donc de prédiction d'un échec thérapeutique un an après l'IRM de fusion étaient moindres pour tous les marqueurs par rapport au sous-groupe précédent (Annexe 9). Les intervalles de confiance des AUC étaient plus larges et comprenaient tous 0,5.

Le seuil  $\geq 3$  nouvelles lésions FUSION avait une sensibilité de 25 % et une spécificité de 88,1 % avec une AUC du nombre de lésions FUSION à 0,624 [0,469-0,779] et une exactitude à 75,9 %. Les AUC PIXYL ne différaient pas significativement de celle du nombre de nouvelles lésions FUSION mais étaient toutes inférieures. Elles variaient de 0,544 pour le ratio des volumes de nouvelles lésions à 0,615 pour le volume évolutif. L'AUC du nombre de nouvelles lésions PIXYL de taille seuil 3 mm était de 0,574 [0,413 - 0,734]. De meilleures sensibilités PIXYL étaient toujours contrebalancées par des spécificités moindres à la FUSION de manière significative. C'était le cas du nombre de nouvelles lésions PIXYL  $\geq 3$  avec une sensibilité de 75 % (vs 25 % pour la FUSION ;  $p = 0,01$ ) et une spécificité de 50,7 % (vs 88,1 % ;  $p < 0,001$ ). Le seuil  $\geq 3$  nouvelles lésions PIXYL de taille seuil 3 mm avait une sensibilité de 62,5 % comparable à FUSION ( $p = 0,07$ ) mais une spécificité moindre de 58,2 % ( $p < 0,001$ ) de même que l'exactitude de 59 % ( $p = 0,03$ ). Les marqueurs PIXYL ayant leur meilleur seuil sensibilité + spécificité similaire à FUSION ( $p > 0,07$ ) étaient seulement les volumes ( $\geq 417$  mm<sup>3</sup>) et ratio ( $\geq 311$ ) des lésions élargies. Leurs sensibilités étaient respectivement de 56,3 % et 37,5 % et leurs spécificités de 74,6 % et 86,6 %. Le ratio volume / lésion élargie  $\geq 311$  avait par ailleurs la meilleure des exactitudes PIXYL à 77,1 % comparable à celle de 75,9 % pour la référence  $\geq 3$  lésions FUSION ( $p = 1$ ).

Les marqueurs PIXYL n'avaient donc pas de performance globale meilleure que la référence FUSION  $\geq 3$  nouvelles lésions pour prédire un échec thérapeutique postérieur à l'IRM de FUSION. Ceux ayant le profil de performances sensibilité / spécificité le plus proche concernaient les volumes des lésions élargies (ou leur ratio). Le nombre de nouvelles lésions PIXYL avait une spécificité moindre que FUSION. Le meilleur moyen d'augmenter légèrement la sensibilité sans perdre de spécificité était, comme dans le sous-groupe précédent, de fixer un seuil  $\geq 2$  nouvelles lésions pour FUSION (pas de différence significative néanmoins, sensibilité de 50 %, spécificité de 85,1 % et exactitude de 78,3 %).

#### **IV. Recherche de seuils lésionnels PIXYL optimaux**

Afin d'identifier si des seuils lésionnels sélectifs pouvaient améliorer la pertinence clinique des marqueurs PIXYL, des filtres ont été appliqués sur les dimensions dans les trois axes spatiaux x,y,z, les volumes et le degré de certitude des lésions segmentées par PIXYL. Ces filtres ont permis d'obtenir les seuils optimaux pour considérer les lésions selon la certitude, la taille et le volume de leur segmentation PIXYL par rapport à leur nombre (Tableau 10) ou leur volume (Tableau 11) pour identifier les patients avec échec thérapeutique (global) selon les lésions considérées.

La Figure 9 représente la variation de l'AUC du nombre de lésions PIXYL selon le seuil étudié. Les courbes pour les lésions élargies étaient au-dessus de celles des nouvelles lésions en dehors d'un pic à 6 mm pour la taille des nouvelles lésions en y. Une taille seuil  $\geq 6$  mm selon l'axe y et une certitude  $\geq 0,8$  pour les nouvelles lésions PIXYL se distinguaient par un pic notable sur ces courbes.

Les AUC supérieures à celle du nombre de lésions FUSION (0,668) ont été mises en gras dans les tableaux Tableau 10 et Tableau 11 mais aucune ne différait significativement (soulignage). La combinaison des seuils filtres optimaux pour chaque critère ne permettait pas d'obtenir de meilleures performances selon l'AUC que le meilleur seuil filtre unique basé sur la taille selon l'axe y  $\geq 6$  mm pour les nouvelles lésions avec un seuil discriminant  $\geq 1$  nouvelle lésion répondant au critère de taille considérée. Ce seuil PIXYL avait une AUC de 0,707 [0,599 - 0,815]. Il avait une meilleure sensibilité à 70 % par rapport aux 30,3 % de celle  $\geq 3$  lésions FUSION ( $p = 0,003$ ) mais une spécificité moindre à 70 % versus 96 % de FUSION

( $p < 0,001$ ). Son exactitude de 69,9 % était exactement identique à FUSION ( $p = 1$ ). Le nombre de lésions PIXYL filtré avec cette taille seuil  $y \geq 6$  mm semblait un peu mieux corrélé avec le nombre de lésions FUSION (Tau de Kendall à 0,38 ;  $p < 0,001$ ) que le compte de nouvelles lésions PIXYL sans (0,24) et avec taille seuil 3D de 3 mm (0,26) même si cela restait faible.

Variable	Seuil de la variable	AUC [IC95%]	Nombre seuil de lésions optimal	Se	Spé
Certitude nouvelle lésion	0,8	0,652 [0,543-0,761]	1	<b>0.55</b>	<u>0.74</u>
Certitude lésion élargie	0,7	0,665 [0,552-0,777]	1	<b>0.64</b>	<u>0.7</u>
Taille nouvelle lésion en xyz	6	0,641 [0,54-0,742]	1	<b>0.42</b>	0.84
Taille lésion élargie en xyz	7	0,656 [0,548-0,763]	1	<b>0.55</b>	<u>0.78</u>
Taille nouvelle lésion en x	5	0,625 [0,505-0,744]	2	<b>0.42</b>	0.82
Taille lésion élargie en x	6	0,654 [0,541-0,767]	1	<b>0.61</b>	<u>0.68</u>
<b>Taille nouvelle lésion en y</b>	6	<b>0,707 [0,599-0,815]</b>	1	<b>0.7</b>	<u>0.7</u>
Taille lésion élargie en y	5	0,661 [0,544-0,778]	2	<b>0.55</b>	<u>0.78</u>
Taille nouvelle lésion en z	5	0,629 [0,512-0,745]	2	<b>0.39</b>	0.82
Taille lésion élargie en z	7	0,651 [0,54-0,762]	1	<b>0.58</b>	<u>0.72</u>
Volume nouvelle lésion	20	0,647 [0,525-0,769]	2	<b>0.58</b>	<u>0.7</u>
Volume lésion élargie	100	0,662 [0,552-0,773]	1	<b>0.61</b>	<u>0.72</u>

Tableau 10. Seuils lésionnels PIXYL optimaux considérés par rapport au compte lésionnel (Tailles en mm et Volumes en mm<sup>3</sup> ; Se : sensibilité ; Spé : spécificité)

Variable	Seuil de la variable	AUC [IC95%]	Volume seuil optimal	Se	Spé
Taille nouvelle lésion en x	3	0.628 [0.502-0.755]	108	<b>0.58</b>	<u>0.68</u>
Taille lésion élargie en x	0	<b>0.676 [0.557-0.795]</b>	181	<b>0.61</b>	<u>0.76</u>
Taille nouvelle lésion en y	6	<b>0.692 [0.582-0.802]</b>	11	<b>0.7</b>	<u>0.7</u>
Taille lésion élargie en y	8	<b>0.678 [0.564-0.791]</b>	488	<b>0.45</b>	0.88
Taille nouvelle lésion en z	3	0.632 [0.504-0.759]	65	<b>0.7</b>	<u>0.58</u>
Taille lésion élargie en z	0	<b>0.676 [0.557-0.795]</b>	181	<b>0.61</b>	<u>0.76</u>
Volume nouvelle lésion	20	0,636 [0,512-0,76]	60	<b>0.64</b>	<u>0.68</u>
Volume lésion élargie	100	<b>0,682 [0,572-0,792]</b>	195	<b>0.58</b>	<u>0.78</u>

Tableau 11. Seuils lésionnels PIXYL optimaux considérés par rapport au volume lésionnel (Tailles en mm et Volumes en mm<sup>3</sup> ; Se : sensibilité ; Spé : spécificité)

L'application de la taille seuil de 6 mm aux 3 axes pour les nouvelles lésions donnait des performances significativement comparables à la référence de 3 lésions FUSION. L'AUC était de 0,641 [0,54 - 0,742]. La sensibilité était de 42 % (vs 30,3 % ;  $p = 0,44$ ) et la spécificité de 84 % (vs 96 % ;  $p = 0,09$ ) pour le seuil optimal  $\geq 1$  nouvelle lésion. Enfin l'exactitude était de 67,5 % pour ce seuil (vs 69,9 % ;  $p = 0,87$ ).

La recherche de seuils lésionnels optimaux dans les deux sous-groupes donnait des résultats semblables (résultats dans les Annexes (Annexe 10 à Annexe 13). Des seuils proches permettaient de filtrer les lésions plus spécifiques. Les AUC ne différaient pas significativement de la référence du nombre de nouvelles lésions FUSION. Les paramètres de sensibilité / spécificité pour les meilleurs seuils étaient soit statistiquement proches de la FUSION soit inversés avec une augmentation de la sensibilité aux dépens d'une perte inacceptable de la spécificité.

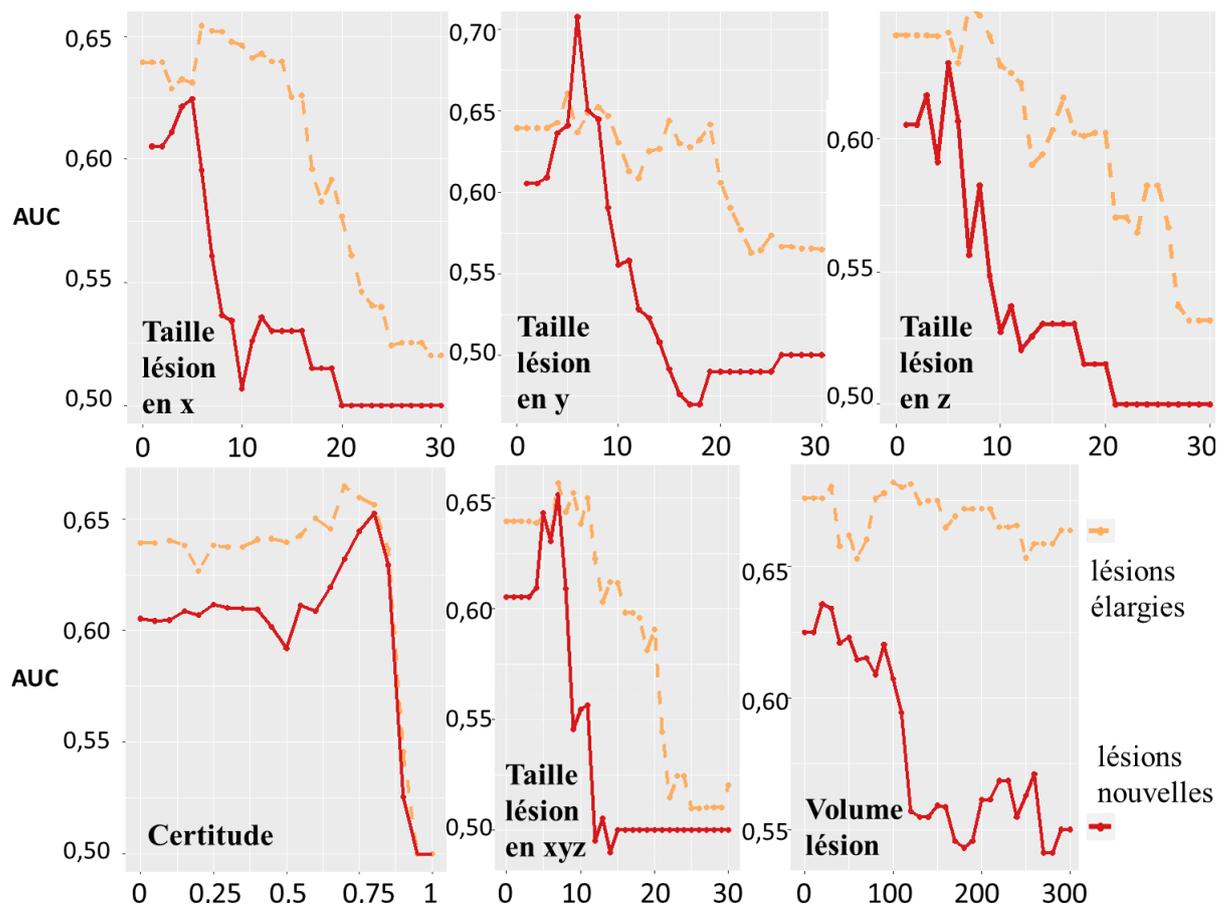


Figure 9. Courbes des AUC du nombre de lésions selon les principaux seuils PIXYL pour les lésions nouvelles (courbes rouges) et élargies (courbe orange en pointillé) (seuil en mm pour les tailles de lésions, certitude de 0 à 1 et volume en mm<sup>3</sup>)

La recherche de seuils lésionnels concernant les dimensions, les volumes et le degré de certitude des lésions segmentées par PIXYL par rapport au nombre et au volume des lésions n'a pas permis de mettre en évidence de performances significativement meilleures que la FUSION pour identifier un échec thérapeutique. La taille seuil optimale d'une nouvelle lésion PIXYL pour identifier l'échec thérapeutique était de 6 mm.

## **V. Combinaison de marqueurs radiologiques dans des modèles de régression logistique d'identification de l'échec thérapeutique global**

Une segmentation automatique avec PIXYL apportant des informations supplémentaires par rapport à une interprétation radiologique courante, nous avons cherché si la combinaison pondérée des différentes données concernant le nombre et le volume des lésions nouvelles et élargies permettait d'améliorer les performances pour identifier l'échec thérapeutique global. Et dans un second temps, bien que les résultats précédents n'aient pas montré de supériorité significative d'une méthode sur l'autre, chacune avait son point fort : une forte spécificité mais un manque de sensibilité pour la FUSION et inversement pour PIXYL. Nous avons donc cherché à construire un modèle statistique combinant les renseignements fournis par les deux méthodes.

Différents modèles de régression logistique de type binomiale ont été construits et évalués selon l'AUC sensibilité / spécificité par rapport à la référence obtenue avec FUSION tout en essayant d'obtenir la meilleure sensibilité du seuil optimal en conservant une spécificité d'au moins 90 % jugée indispensable pour une application clinique. Les meilleurs modèles obtenus avec un grand nombre de variables concernant les nombres, volumes, ratio et certitude de la segmentation pour différentes tailles seuil étaient ensuite simplifiés en essayant de retirer la variable la moins utile tout en gardant de bonnes performances. Le modèle de régression logistique binomiale de référence construit avec le nombre de lésions FUSION avait un odds ratio significatif de 2,02 [1,34 - 3,29] ( $p = 0,002$ ). Pour une nouvelle lésion FUSION, le risque de présenter un échec thérapeutique était donc doublé (+ 102 %). Pour deux lésions FUSION, le risque était multiplié par 4,08 et pour trois lésions FUSION, il était 8,24 fois plus important que sans lésion FUSION. L'odds ratio obtenu en univarié pour une nouvelle lésion PIXYL était

de 1,07 (1,01 - 1,16 ;  $p = 0,044$ ) et de 1,10 (1,01 - 1,22 ;  $p = 0,037$ ) si taille seuil de 3 mm. Il était donc significativement moindre que pour une nouvelle lésion FUSION (intervalles de confiance non chevauchant). La probabilité d'échec augmentait de 22,5 % et 33,1 % au seuil  $\geq 3$  nouvelles lésions PIXYL sans et avec taille seuil de 3 mm et de 61 % pour le seuil  $\geq 5$  nouvelles lésions PIXYL  $\geq 3$  mm.

Le meilleur modèle sélectionné appelé *Modèle PIXYL* prenait en compte le nombre de nouvelles lésions pour chaque territoire (infratentoriel, profond, périventriculaire et juxtacortical), le degré de certitude de la segmentation pour les lésions nouvelles et élargies de manière séparée, le nombre de lésions élargies et les volumes des nouvelles lésions et des lésions élargies de manière distincte. Le nombre de lésions infratentorielles était le seul compte de nouvelles lésions nécessaire dans le modèle simplifié mais néanmoins les simplifications des modèles résultaient toujours d'une perte de spécificité jugée incompatible avec une application clinique. La courbe AUC de ce modèle a été dessinée dans la [Figure 10](#) et était de 0,758 [IC95% 0,651 - 0,866] ( $p = 0,25$  par rapport à FUSION 0,668). Le seuil ayant le meilleur rapport sensibilité / spécificité était que la probabilité d'un échec soit estimée  $\geq 0,499$  avec une sensibilité de 54,5 % ( $p = 0,08$  par rapport à  $\geq 3$  lésions FUSION 30,3 %) et une spécificité de 90 % ( $p = 0,44$  par rapport à  $\geq 3$  lésions FUSION 96 %). L'exactitude de ce modèle était de 75,9 % similaire aux 69,9 % de la référence FUSION ( $p = 0,49$ ). Néanmoins, la validité de ce modèle était à interpréter avec précaution, les odds ratio n'étant pas significatifs en dehors de celui pour le nombre de nouvelles lésions infratentorielles qui était de 3,42 [1,27 - 11,47] en univarié ( $p = 0,03$ ) et 4,61 [1,39 - 19,51] en multivarié ( $p = 0,002$ ) ([Tableau 12](#)). **La présence de lésions infratentorielles PIXYL était donc un marqueur de risque significatif de présenter un échec thérapeutique dont l'odds ratio semblait plus important que pour les lésions FUSION mais était statistiquement similaire (chevauchement des intervalles de confiance).**

L'ajout de la donnée 'nombre de nouvelles lésions FUSION' a permis d'améliorer ce modèle sous le nom de *Modèle GLOBAL* avec la meilleure AUC possible à 0,836 [IC95% 0,740 - 0,931]. L'AUC était même significativement supérieure à celle de la FUSION seule (0,668) ( $p = 0,003$ ) ([Figure 10](#)). En effet au seuil de probabilité  $\geq 0,548$ , pour une spécificité identique au 'nombre de lésions FUSION  $\geq 3$ ' à 96 %, la sensibilité était plus élevée à 63,6 % versus 30,3 % ( $p = 0,01$ ). Le doublement de sensibilité du modèle permettait donc un gain de 33 % (+ 11 / 33 patients avec échec) de la sensibilité par rapport à la référence d'au moins trois nouvelles lésions FUSION et de 21 % (+ 7 / 33 patients avec échec) par rapport au seuil

FUSION optimal de deux nouvelles lésions. L'exactitude était de 83,1 % soit similaire aux 69,9 % du seuil de référence  $\geq 3$  lésions FUSION ( $p = 0,07$ ). En multivarié, les odds ratio étaient significatifs pour le degré de certitude des nouvelles lésions (29,47), le nombre de nouvelles lésions infratentorielles (7,68), le nombre de lésions élargies (0,63) et le nombre de nouvelles lésions FUSION (3,03) (Tableau 12). En univarié, les odds ratio n'étaient plus significatifs pour le degré de certitude des nouvelles lésions et le nombre de lésions élargies. L'odds ratio de 0,63 [0,38 - 0,89] pour le nombre de lésions élargies traduisait une association négative puisqu'une lésion élargie identifiée PIXYL diminuait de 37 % la survenue d'un échec thérapeutique.

Les marqueurs radiologiques significatifs dans le *Modèle GLOBAL* ont amené la construction d'un modèle les regroupant sous le nom de *Modèle GLOBAL simplifié* en excluant le nombre de lésions élargies PIXYL en raison d'un odds ratio inattendu inférieur à 1. De plus, l'exclusion de cette dernière variable ne modifiait pas les résultats. Ce dernier modèle était statistiquement aussi performant que les deux premiers avec une AUC de 0,783 [IC95% 0,678 - 0,888] meilleure que celle du « nombre de lésions FUSION  $\geq 3$  » ( $p = 0,02$ ). Il restait pertinent avec une meilleure sensibilité que la FUSION à 68,8 % ( $p = 0,046$ ) malgré une spécificité inférieure mais statistiquement proche à 82,1 % ( $p = 0,68$ ) (Figure 10). L'exactitude était de 78,3 % ( $p = 0,29$ ). L'odds ratio du modèle multivarié pour le degré de certitude des nouvelles lésions perdait la significativité statistique par rapport au modèle précédent. Les nombres de nouvelles lésions infratentorielles PIXYL et de nouvelles lésions FUSION restaient significatifs avec des odds ratio respectivement à 6,49 [1,99 - 26,87] et 2,28 [1,44 - 3,98] (Tableau 12).

Le nombre global de nouvelles lésions PIXYL n'était pas aussi intéressant que le nombre de nouvelles lésions par territoire, notamment pour le territoire infratentoriel, dans la construction des modèles. L'utilisation des seuils ayant le meilleur rapport sensibilité + spécificité pour les différents marqueurs radiologiques afin de dichotomiser ces marqueurs en variables binaires ne permettait pas d'obtenir de meilleurs modèles de régression logistique que les modèles construits sur les variables quantitatives, la spécificité semblant souvent moindre.

Des modèles de régression logistique incluant les seuils optimaux identifiés dans la partie IV. Recherche de seuils lésionnels PIXYL optimaux ont été générés. Pour la dimension  $y \geq 6$  mm, l'odds ratio était de 1,56 (1,16 - 2,26,  $p = 0,008$ ) pour une nouvelle lésion PIXYL en

univarié. Un modèle basé sur les dimensions seuil optimales en y (6 mm pour les nouvelles lésions et 5 mm pour les lésions élargies) a été construit avec les comptes et volumes des lésions PIXYL nouvelles et élargies. L'AUC était de 0,691 [0,567 – 0,814]. Le seuil de probabilité optimale  $\geq 0,327$  avait une sensibilité de 63,6 % et une spécificité de 78 %. Le profil de la courbe ROC semblait moins approprié à un usage clinique que les modèles précédents avec une sensibilité initiale de 21,2 % pour une spécificité de 94 % au seuil de probabilité de 0,65. L'ajout du nombre de lésions FUSION à ce modèle donnait une AUC de 0,748 [0,637-0,859]. Au seuil optimal de probabilité d'échec de 0,647, la sensibilité était de 45,5 % et la spécificité de 98 %. Ces valeurs semblaient donc plus proches de celles du modèle FUSION au seuil  $\geq 2$  lésions que des meilleurs modèles combinant des données PIXYL et FUSION présentés ci-dessus.

<b><i>Modèle PIXYL</i></b>	<b>OR (univarié) [IC95%]</b>	<b><i>p</i></b>	<b>OR (multivarié) [IC95%]</b>	<b><i>p</i></b>
Correction (exp(intercept $\beta_0$ ))			0,2 [0,01-3,32]	0,267
Nombre de nouvelles lésions profondes	1,24 [0,91-1,75]	0,188	1,04 [0,64-1,64]	0,857
Nombre de nouvelles lésions infratentorielles	3,42 [1,27-11,47]	<b>0,028</b>	4,61 [1,39-19,51]	<b>0,002</b>
Nombre de nouvelles lésions juxtacorticales	1,09 [0,99-1,21]	0,095	0,98 [0,79-1,20]	0,854
Nombre de nouvelles lésions périvertriculaires	1,16 [0,93-1,47]	0,198	0,96 [0,68-1,37]	0,831
Certitude nouvelle lésion	2,89 [0,34-27,12]	0,337	12,21 [0,93-223,61]	0,07
Nombre de lésions élargies	1,06 [0,92-1,23]	0,433	0,76 [0,51-1,03]	0,118
Certitude lésion élargie	0,51 [0,06-4,42]	0,538	0,31 [0,02-3,99]	0,368
Volume nouvelle lésion	1 [1,00-1,00]	<b>0,044</b>	1 [1,00-1,01]	0,167
Volume lésion élargie	1 [1,00-1,00]	0,06	1 [1,00-1,00]	0,152
<b><i>Modèle GLOBAL</i></b>				
Correction (exp(intercept $\beta_0$ ))			0,095 [0,004-2,02]	0,139
Nombre de nouvelles lésions profondes	1,24 [0,91-1,75]	0,188	1,02 [0,59-1,67]	0,937
Nombre de nouvelles lésions infratentorielles	3,42 [1,27-11,47]	<b>0,028</b>	7,68 [1,85-47,49]	<b>0,014</b>
Nombre de nouvelles lésions juxtacorticales	1,09 [0,99-1,21]	0,095	1,19 [0,92-1,56]	0,184
Nombre de nouvelles lésions périvertriculaires	1,16 [0,93-1,47]	0,198	1 [0,67-1,51]	0,995
Certitude nouvelle lésion	2,89 [0,34-27,12]	0,337	29,47 [1,49-957,38]	<b>0,038</b>
Nombre de lésions élargies	1,06 [0,92-1,23]	0,433	0,63 [0,38-0,89]	<b>0,023</b>
Certitude lésion élargie	0,51 [0,06-4,42]	0,538	0,18 [0,01-2,57]	0,205
Volume nouvelle lésion	1 [1,00-1,00]	<b>0,044</b>	1 [1,00-1,00]	0,94
Volume lésion élargie	1 [1,00-1,00]	0,06	1 [1,00-1,00]	0,339
Nombre de lésions FUSION	2,02 [1,34-3,29]	<b>0,002</b>	3,03 [1,58-6,99]	<b>0,003</b>
<b><i>Modèle GLOBAL simplifié</i></b>				
Correction (exp(intercept $\beta_0$ ))			0,04 [0,004-0,293]	<b>0,003</b>
Nombre de nouvelles lésions infratentorielles	3,42 [1,27-11,47]	<b>0,028</b>	6,49 [1,99-26,87]	<b>0,004</b>
Certitude nouvelle lésion	2,89 [0,34-27,12]	0,337	14,78 [1,03-274,46]	0,056
Nombre de lésions FUSION	2,02 [1,34-3,29]	<b>0,002</b>	2,28 [1,44-3,98]	<b>0,001</b>

Tableau 12. Odds ratio (OR) pour les variables des trois meilleurs modèles de régression logistique sélectionnés

La formule d'un modèle de régression logistique est : probabilité (x) =  $\exp(\beta_0 + \beta_1x + \beta_2x + \dots + \beta_px) / (1 + \exp(\beta_0 + \beta_1x + \beta_2x + \dots + \beta_px))$  où  $\beta$  désignent les coefficients pondérant les variables incluses,  $\beta_0$  étant la correction du modèle pour  $x=0$ , et  $x$  la valeur de la variable considérée (par exemple  $\beta_1$  où  $x$  est le compte de lésions). L'exponentielle des coefficients donne les odds ratio correspondants.

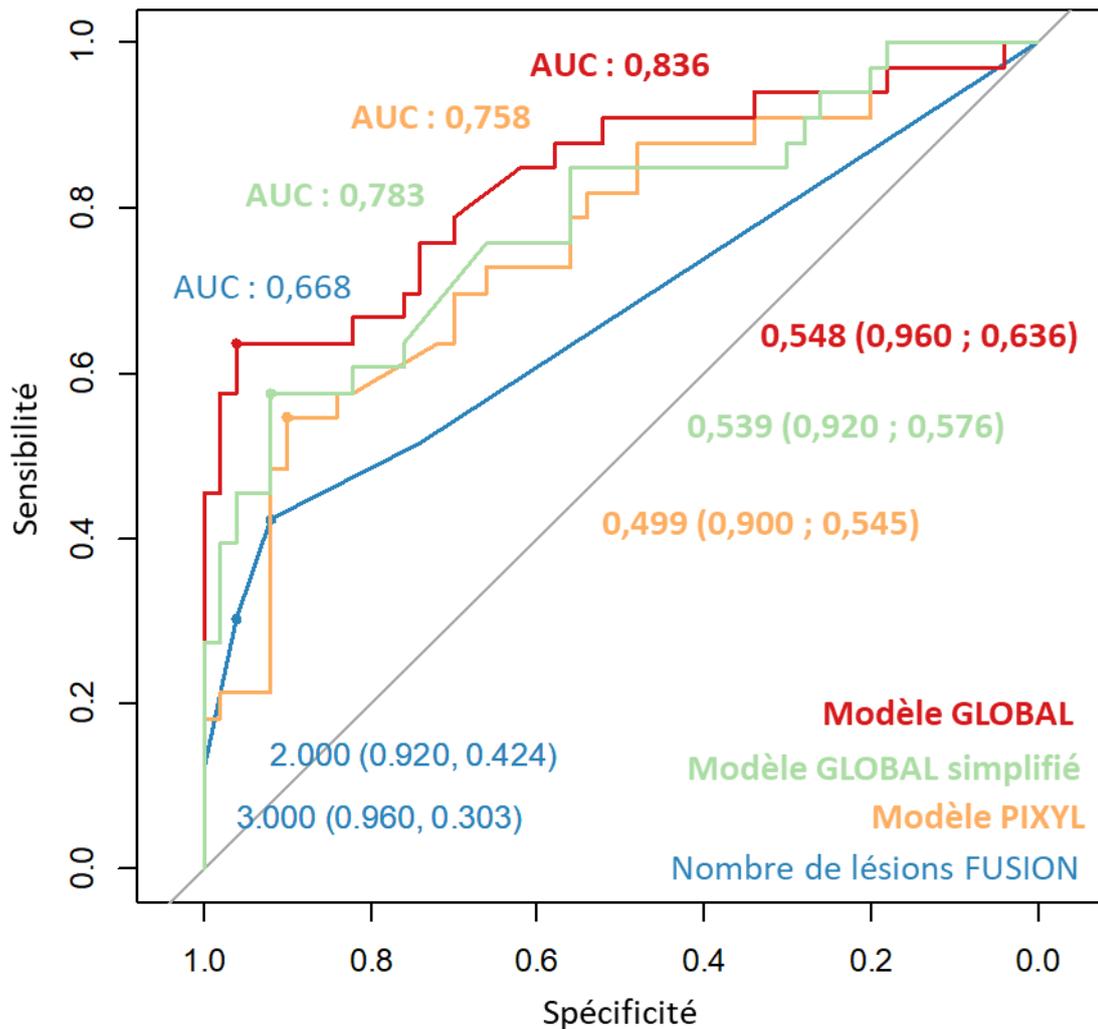


Figure 10. Courbes sensibilité / spécificité des modèles de régression logistique comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe du modèle et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité)

La génération de modèles de régression logistique dans les deux sous-groupes d'échec thérapeutique n'a pas permis d'obtenir de gains de performance significatifs quant à l'identification ou la prédiction de la survenue d'un échec thérapeutique (résultats non présentés). Les performances étaient aux mieux comparables au seul nombre de nouvelles lésions FUSION. De plus la réduction du nombre d'échecs observés par rapport au groupe global rendait les intervalles de confiance plus larges et les comparaisons moins fiables.

Les modèles de régression logistique combinant différents marqueurs PIXYL avaient au mieux des performances comparables à celles du nombre de lésions FUSION. L'association de marqueurs PIXYL et FUSION permettait d'obtenir une AUC (0,836) et une sensibilité (64 %) meilleures que le seul nombre de lésions FUSION au seuil de référence de 3. Ce *Modèle GLOBAL* améliorait l'AUC de 25 % et doublait la sensibilité par rapport à la référence FUSION, permettant d'identifier 33 % de patients supplémentaires avec échec thérapeutique pour une spécificité conservée (96 %). Les marqueurs PIXYL les plus utiles dans ces modèles étaient le nombre de nouvelles lésions infratentorielles et le degré de certitude de segmentation des nouvelles lésions.

Les seuils de taille ne permettaient pas de créer un modèle significativement meilleur que l'absence de seuil filtre pour les lésions.

La construction de modèles multivariés dans les deux sous-groupes d'échec thérapeutique n'était pas concluante.

## **VI. Analyses de concordance des marqueurs PIXYL avec la FUSION et l'échec thérapeutique global.**

Dans les résultats présentés en partie II, il n'existait pas de corrélation forte entre le nombre de nouvelles lésions PIXYL et FUSION (Tau de Kendall  $< 0,5$ ). Avoir une nouvelle lésion PIXYL d'au moins 3 mm permettait d'identifier 27 / 30 patients (sensibilité de 90 %) avec une nouvelle lésions FUSION. Cependant, 40 / 67 patients (59,7 %) ayant au moins une lésion PIXYL d'au moins 3 mm n'avaient pas de lésion FUSION (précision VPP de 40,3 %), soit une spécificité de 24,5 % (13 / 53). Nous avons testé la concordance des seuils  $\geq 3$  nouvelles lésions entre les deux méthodes et avec le critère de taille seuil  $\geq 3$  mm [Figure 11](#). Les coefficients Kappa de Cohen à 0,16 [0,03 – 0,31] et 0,23 [0,08 – 0,41] dénotaient une mauvaise concordance en accord avec la grande dispersion du nombre de lésions PIXYL. La majorité des patients avec au moins trois nouvelles lésions FUSION était bien identifiée avec le seuil de trois nouvelles lésions PIXYL  $\geq 3$  mm (10 / 12 ; 83,3 %). Mais 28 patients supplémentaires répondaient à ce seuil PIXYL sans répondre au seuil FUSION, soit 73,7 % de faux positifs et une spécificité de 60,6 % (43 / 71). Le seuil optimal  $\geq 5$  nouvelles lésions PIXYL  $\geq 3$  mm ne permettait pas de trouver une meilleure concordance avec le seuil de référence FUSION (Kappa

de Cohen à 0,35 [0,12 - 0,58]). Les concordances entre l'échec thérapeutique global observé et ces seuils  $\geq 3$  nouvelles lésions PIXYL sans et avec taille seuil  $\geq 3$  mm n'étaient pas meilleures et semblaient même moindres (Kappa de Cohen respectivement à 0,19 [-0,01 - 0,39] et 0,19 [-0,02 - 0,4]). La concordance du seuil  $\geq 3$  nouvelles lésions FUSION avec l'échec thérapeutique n'était pas non plus très bonne (Kappa de Cohen à 0,29 [0,12 - 0,48]) même pour le seuil  $\geq 2$  (0,37 [0,18 - 0,56]).

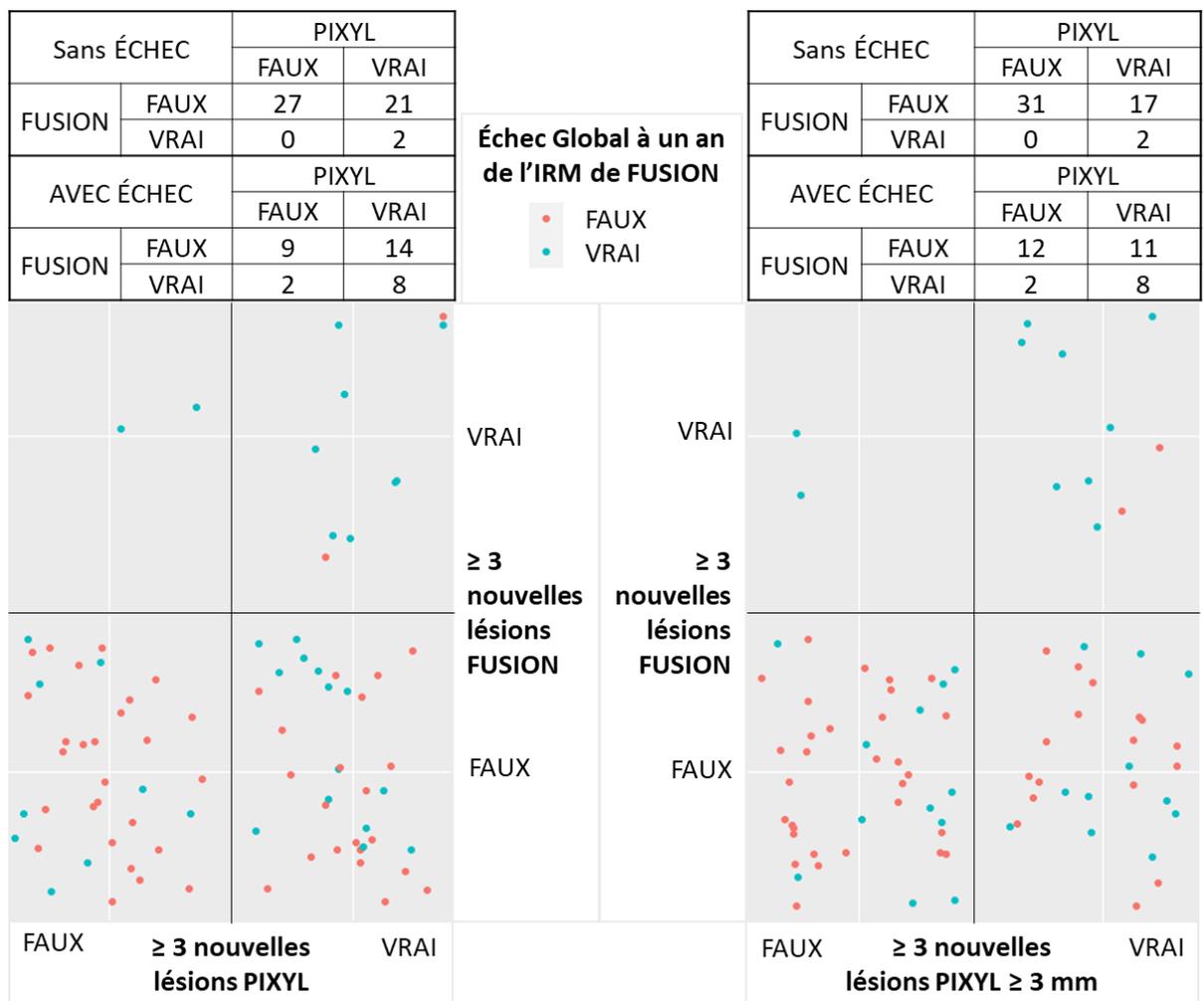


Figure 11. Analyses de concordance des seuils  $\geq 3$  nouvelles lésions entre PIXYL et FUSION

La concordance entre le seuil  $\geq 3$  nouvelles lésions FUSION et l'échec entre les deux IRM restait faible bien que semblant meilleure que pour l'échec global (Tau de Kendall de 0,43 [0,21 - 0,64] ;  $p < 0,001$ ). Ce seuil  $\geq 3$  nouvelles lésions PIXYL avait une concordance faible et non significative de 0,16 [-0,03 - 0,34] ( $p = 0,09$ ) sans taille seuil et de 0,18 [-0,02 - 0,38] ( $p = 0,09$ ) pour la taille seuil de 3 mm.

La concordance entre le seuil  $\geq 3$  nouvelles lésions FUSION et l'échec après l'IRM de FUSION était très faible et non significative (Tau de Kendall de 0,14 [-0,07 - 0,43] ;  $p = 0,25$ ). Il en était de même pour PIXYL sans taille seuil (0,16 [-0,02 - 0,35] ;  $p = 0,054$ ) et avec taille seuil de 3 mm (0,14 [-0,04 - 0,32] ;  $p = 0,14$ ).

La concordance entre le *Modèle PIXYL* pour une probabilité d'échec prédit à la valeur seuil  $\geq 0,499$  et  $\geq 3$  nouvelles lésions FUSION était mauvaise avec un Kappa de Cohen à 0,33 [0,11 - 0,56] (Figure 12). Cette concordance était logiquement correcte mais sans être excellente (Kappa de Cohen à 0,47 [0,25 - 0,69]) pour le *Modèle GLOBAL* avec un seuil de probabilité  $\geq 0,548$ . La Figure 12 montre que les deux faux positifs du *Modèle GLOBAL* n'étaient pas les mêmes que ceux de la FUSION qui ont été discutés dans la partie III. L'un d'eux a eu une progression confirmée de l'EDSS un peu plus de deux ans après l'IRM de FUSION et n'avait qu'une nouvelle lésion comptée en FUSION alors que l'autre n'a eu aucune activité clinique comme radiologique dans les deux ans de suivi disponibles suite à l'IRM de FUSION. La concordance entre l'échec thérapeutique global et les seuils de probabilité optimaux des modèles des modèles de régression logistique était moyenne pour le *Modèle PIXYL* (Kappa de Cohen à 0,47 [0,26 - 0,66]) et bonne pour le *Modèle GLOBAL* (Kappa de Cohen à 0,63 [0,44 - 0,78]).

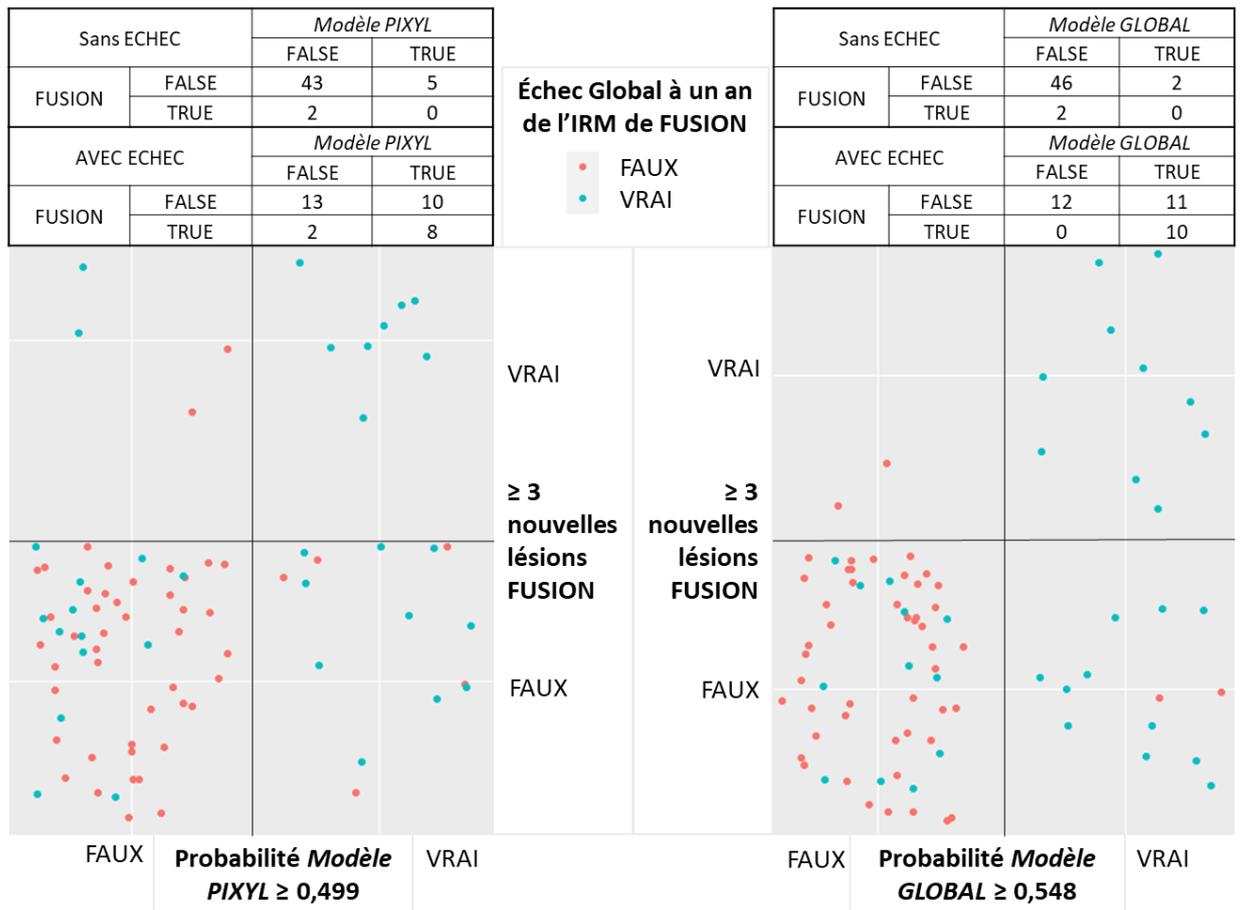


Figure 12. Analyses de concordance des seuils de probabilité optimaux entre les modèles de régression logistique sélectionnés et  $\geq 3$  nouvelles lésions FUSION

Les concordances binaires entre les seuils de nouvelles lésions PIXYL et FUSION étaient faibles. Elles étaient également faibles entre ces seuils et la survenue d'un échec thérapeutique. Les modèles de régression logistique avaient une concordance plus correcte avec la survenue d'un échec thérapeutique. Cette concordance était bonne pour le *Modèle GLOBAL* au seuil de probabilité optimal  $\geq 0,548$  (Kappa de Cohen à 0,63).

## VII. Vitesse de survenue de l'échec thérapeutique.

La vitesse de survenue de l'échec thérapeutique dans l'année suivant l'IRM de FUSION selon la classification effectuée par les seuils  $\geq 3$  nouvelles lésions FUSION ou PIXYL ou bien les modèles de régression logistique a été illustrée par les courbes de survie avec test du logrank dans la [Figure 13](#). Le groupe des 12 patients  $\geq 3$  nouvelles lésions FUSION ne subissait pas plus rapidement un échec thérapeutique ( $p = 0,095$ ) mais le seuil optimal  $\geq 2$  nouvelles lésions FUSION retrouvé dans les analyses précédentes permettait de distinguer 18 patients subissant plus rapidement un échec thérapeutique ( $p < 0,001$ ).

Les marqueurs PIXYL ne permettaient pas de bien séparer les patients subissant rapidement un échec des patients stables. Le nombre de nouvelles lésions  $\geq 3$  n'était pas significatif ( $p = 0,06$  et  $p = 0,13$  si taille seuil  $\geq 3$  mm) même pour le seuil  $\geq 5$  lésions de taille  $\geq 3$  mm ( $p = 0,18$ ). Le nombre  $\geq 3$  lésions élargies sans taille seuil et  $\geq 3$  mm n'était pas discriminant non plus ( $p = 0,24$ ). Le volume des nouvelles lésions ne permettait pas de bien distinguer les patients présentant rapidement un échec thérapeutique que ce soit avec les seuils de 106 ou 78 mm<sup>3</sup> optimaux pour l'échec global ou après l'IRM de FUSION ( $p = 0,49$  et  $p = 0,55$ ). Le volume des lésions élargies était un marqueur PIXYL intéressant pour un seuil  $\geq 417$  mm<sup>3</sup> ( $p = 0,007$ ) et inclus dans le ratio volume lésionnel évolutif / nombre de lésions évolutives avec un seuil  $\geq 66$  ( $p = 0,03$ ).

Les modèles de régression logistique *Modèle PIXYL* et *Modèle GLOBAL* obtenaient de bons résultats pour séparer les patients présentant rapidement un échec thérapeutique ( $p < 0,001$  pour ces deux modèles) ([Figure 13](#)).

Parmi les marqueurs radiologiques isolés, le seuil  $\geq 2$  nouvelles FUSION permettait donc de distinguer avec la meilleure significativité les patients ayant le plus rapidement un échec thérapeutique dans l'année suivant l'IRM de FUSION contrairement au seuil de référence  $\geq 3$ .

Peu de marqueurs PIXYL (même avec leurs seuils optimaux) distinguaient les patients pour lesquels l'échec thérapeutique arrivait plus précocement. Le volume des lésions élargies et le ratio du volume lésionnel évolutif étaient les marqueurs PIXYL les plus utiles. Les marqueurs concernant les nouvelles lésions PIXYL n'étaient pas pertinents ici.

Les modèles de régression logistique *Modèle PIXYL* et *Modèle GLOBAL* obtenaient de bonnes performances pour leur seuil optimal.

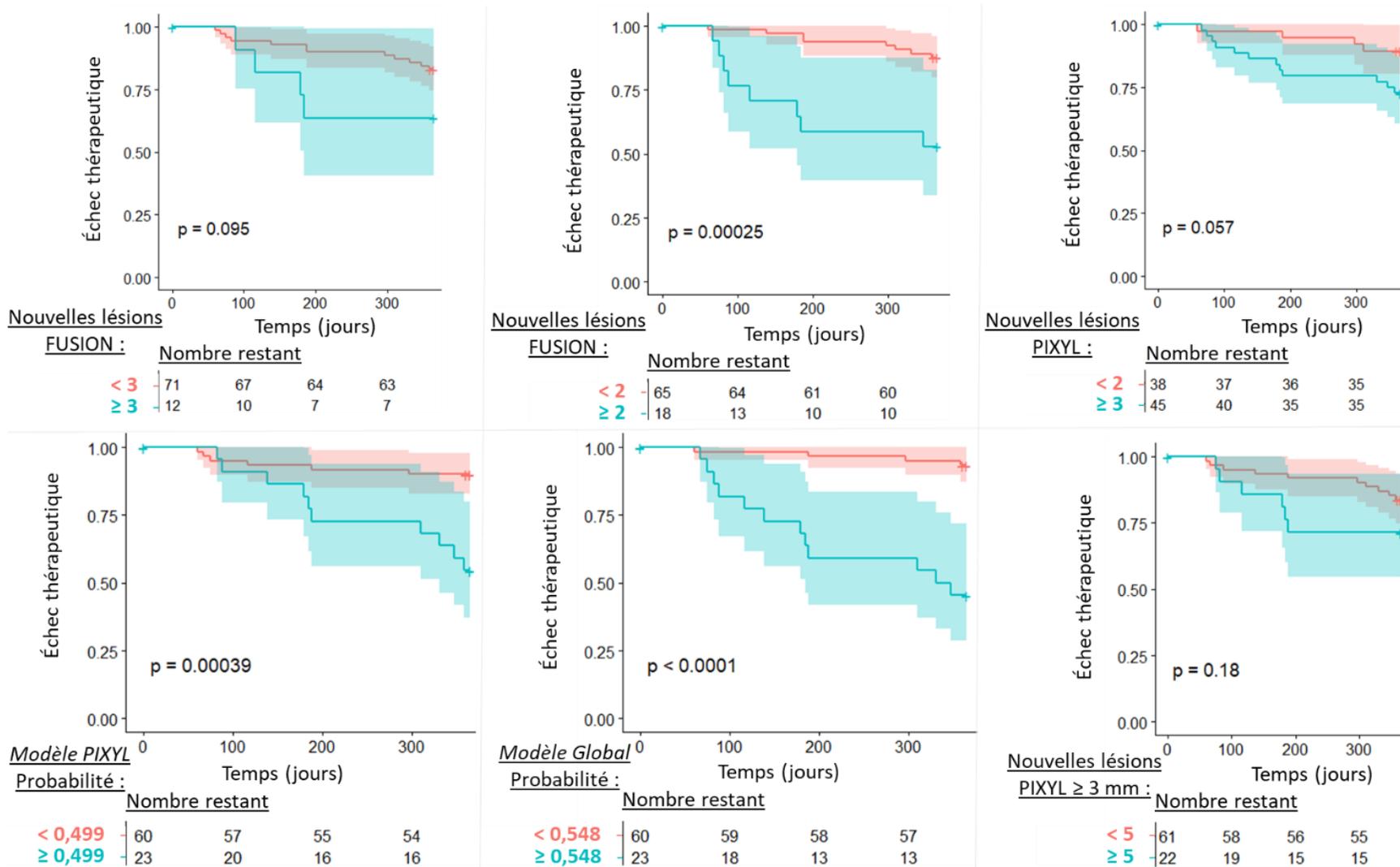


Figure 13. Courbes de survie de la survenue d'un échec thérapeutique dans l'année suivant l'IRM de FUSION et test du logrank

## VIII. Masques de segmentation PIXYL.

Les résultats concernant le nombre de nouvelles lésions PIXYL et notamment la concordance avec la FUSION n'étant pas aussi bonnes qu'espérées, les masques de segmentation PIXYL ont été visualisés pour comprendre à quoi correspondait vraiment les données chiffrées fournies par l'outil de segmentation automatique. Les informations recueillies permettaient de savoir si les nouvelles lésions segmentées étaient validées comme pertinentes, de déterminer les causes des erreurs de segmentation et d'estimer si une segmentation validée avait un intérêt clinique.

Les paramètres d'acquisition FLAIR des IRM comparées n'étaient pas toujours identiques ce qui pouvait induire plus de variations de l'intensité des hypersignaux entre les deux IRM. Notamment, il existait une discordance entre une acquisition avec et sans saturation de graisse chez 36 patients (43,4 %). Étonnamment, le champ d'acquisition FLAIR de l'IRM de FUSION incluait toujours la moelle cervicale haute à l'inverse de l'IRM de référence et quelle que soit les dates des IRM.

La visualisation des masques de segmentation a demandé du temps en raison du nombre important de lésions chez certains patients. Les artefacts faux positifs étaient fréquents, présents chez 63 patients (75,9 %) dont 31 (soit 49,2 % des patients avec artefacts) avaient une IRM parfaitement stable. Les artefacts étaient supposés dus à des variations du contraste de l'intensité de l'hypersignal FLAIR (déjà visible a minima sur l'IRM de référence) chez 47 patients (74,6 % des patients avec artefacts). Ces variations étaient néanmoins détectées à tort de manière variable, difficilement prédictibles et donc peu corrigeables comme illustré par la Figure 14. Les autres artefacts étaient des modifications morphologiques en particulier pour les sillons et les lésions juxtacorticales ou encore les volumes ventriculaires chez 36 (57,1 %), des artefacts de flux du LCR chez 19 (30,2 %) et de potentiels problèmes d'alignement des coupes (la visionneuse ITK n'arrivant pas à les aligner correctement) chez 15 (23,8 %). Par ailleurs, la qualité de l'IRM de référence (rapport signal / bruit et contraste) semblait insuffisante pour 3 patients (3,6 % de la cohorte).

Malgré une sensibilité supposée élevée, neuf patients (10,8 %) comportaient des faux négatifs, c'est-à-dire des nouvelles lésions non détectées comme nouvelles par PIXYL. Il s'agissait souvent d'une erreur de classification en lésion stable ou élargie (7 / 9 ; 77,8 %). La conséquence clinique éventuelle aurait été limitée avec un seul de ces patients pour lequel le seuil de 3 nouvelles lésions FUSION était raté par PIXYL et ce en raison des faux positifs qui

avaient tendance à surestimer les comptes lésionnels également pour les patients ayant des faux négatifs.

Concernant la portée clinique potentielle, 18 segmentations (21,7 %) permettaient de compter des lésions supplémentaires par rapport à FUSION. Le nombre médian de lésions supplémentaires était de  $1,5 \pm 1,75$  [1 - 8]. La plupart de ces lésions étaient douteuses car mal contrastées et de petite taille. La pertinence thérapeutique de la détection de ces lésions supplémentaires n'était pas évidente, c'est-à-dire qu'elles ne modifiaient pas le seuil  $\geq 3$  nouvelles lésions pour 17 d'entre elles. Parmi ces patients, 8 / 18 (44,4 %) ont eu une activité / progression clinique suite à l'IRM de référence jusque dans l'année post-IRM de FUSION. Chez ces 18 patients, quatre avaient déjà une activité radiologique  $\geq 3$  nouvelles lésions FUSION mais 13 n'avaient aucune lésion détectée en FUSION. Un seul de ces patients (1,2 % de la cohorte ou 5,6 % des patients avec détection de lésions non comptées avec FUSION) aurait eu une pertinence thérapeutique potentielle grâce à la détection de trois hypersignaux de taille significative bien que peu intenses et dans une substance blanche « altérée » mais non signalés par FUSION (Figure 15). Dans ce cas particulier, aucun échec thérapeutique n'était signalé. Il s'agissait d'une IRM de rebase line suite à l'introduction d'acétate de glatiramère chez une patiente atteinte d'une forme peu sévère et parfaitement stable au cours du suivi clinique ultérieur.

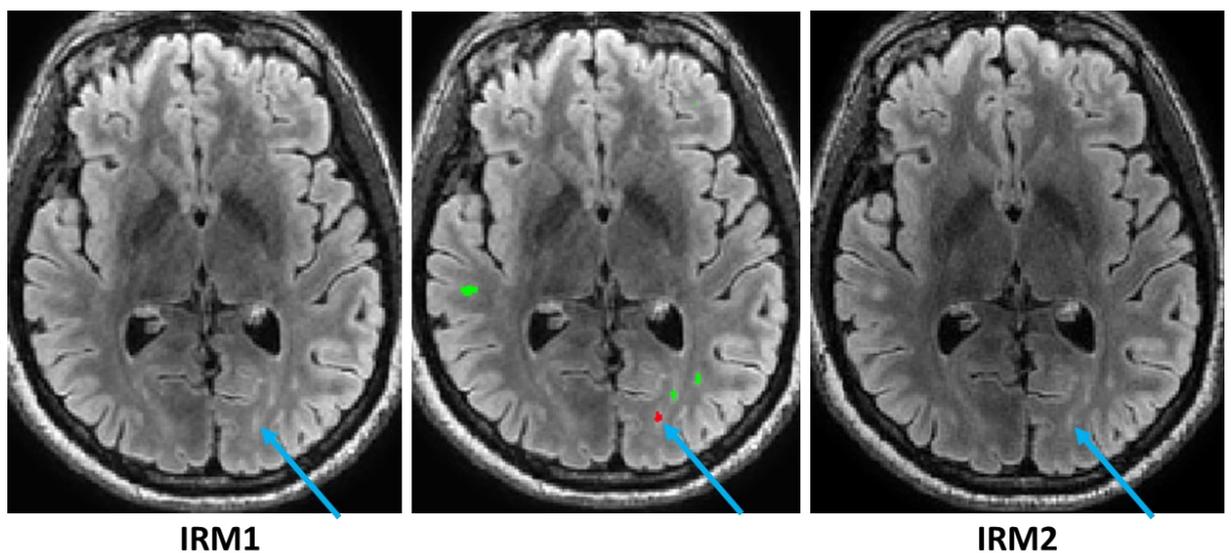


Figure 14. Artefact de contraste typique sur une segmentation PIXYL chez un patient stable malgré des coupes de bonne qualité (artefact désigné par une flèche, à noter la lésion temporale droite plus volumineuse indiquée en vert pour stable a une variation de contraste bien plus marquée mais a bien été segmentée)

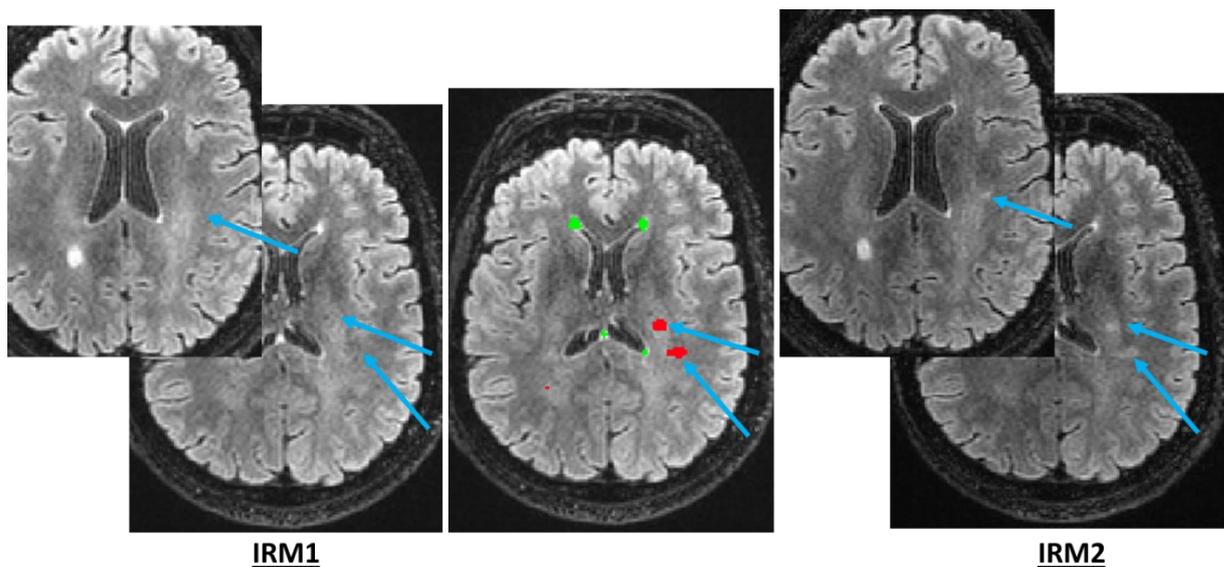


Figure 15. Segmentation PIXYL ayant permis de détecter de nouvelles lésions peu marquées par rapport à la FUSION (en rouge et localisations désignées par des flèches, en vert les lésions stables)

La relecture des masques de segmentation PIXYL a révélé beaucoup d'artefacts faux positifs (75,9 %). Des lésions non détectées par FUSION étaient détectées chez 18 patients. La relecture attentive des masques PIXYL combinée à la FUSION aurait pu classer un patient supplémentaire avec une activité radiologique  $\geq 3$  nouvelles lésions dans un contexte peu intéressant de rebaseline sans échec thérapeutique.

## DISCUSSION

Cette étude était la première évaluation clinique d'une méthode de segmentation automatique pour comparaison d'IRM cérébrales dans le suivi de patients atteints de SEP RR. L'activité clinique et le critère d'échec thérapeutique sont des marqueurs intéressants qui ont été utilisés auparavant pour valider l'intérêt de l'activité radiologique (reposant sur une segmentation manuelle) pour prédire le risque d'évolution sous traitement (47). La plateforme PIXYL, utilisée seule, n'a pas permis d'améliorer la détection des patients avec échec thérapeutique dans l'année suivant l'IRM par rapport à la technique que nous avons choisi comme référence : FUSION. En effet, les performances de classification des meilleurs marqueurs PIXYL, comparées par l'aire sous la courbe ROC sensibilité / spécificité, ne différaient pas significativement de FUSION. Néanmoins, la combinaison de FUSION avec PIXYL a permis la construction de modèles de régression logistique dont les performances pouvaient améliorer celles de FUSION seule. En effet, par rapport au seuil de référence de 'trois nouvelles lésions FUSION', le *Modèle GLOBAL* obtenu améliorait l'AUC de 25 % (de 0,668 à 0,836) et doublait la sensibilité de 30 % à 64 %. Ces performances permettaient d'identifier 33 % de patients supplémentaires avec échec thérapeutique. Cette étude a aussi permis de rapporter l'activité radiologique mesurée par la segmentation automatique PIXYL entre deux IRM cérébrales. Des seuils lésionnels optimaux propres à cette méthode de segmentation ont été déterminés pour identifier les patients avec échec thérapeutique.

Cette étude était basée sur un suivi en vie réelle d'une population représentative de SEP RR ( $n = 83$ ). Notre population était plutôt jeune avec un âge moyen de 39 ans par comparaison à des populations de SEP globales où l'âge moyen est de 51 ans (25). Néanmoins la durée moyenne d'évolution de la maladie de 6 ans montrait que le début de la maladie n'était pas si récent, débutant en moyenne vers 30 ans comme la population d'évaluation de la réponse aux interférons par le score de Sormani (47). L'EDSS médian à 2 témoignait d'un handicap peu élevé pour lequel l'activité inflammatoire est encore prédominante (8) comme dans l'étude de Sormani (47). Les critères d'inclusion reposant sur des patients de la base EDMUS suivis au CHU et la réalisation d'une IRM de FUSION ont sélectionné des patients dont la maladie était assez active. En effet 33 / 83 patients (40 %) ont eu un échec thérapeutique au bout du suivi moyen de deux ans. Dans l'étude de Sormani (47), 23 % des patients avaient un échec à 3 ans, cette proportion variant de 17 à 48 % selon le risque du groupe. Notre effectif était plus limité

avec 83 patients versus 1280 chez Sormani (47). Une activité inflammatoire importante augmentait le nombre d'échec et donc la puissance statistique de notre étude. Les patients cibles étaient donc ceux pour lesquels l'évaluation de l'activité radiologique était la plus utile.

La comparaison des performances d'identification de l'échec thérapeutique des différents marqueurs radiologiques FUSION et PIXYL n'a pas mis en évidence de différence concernant les AUC des meilleures courbes sensibilité / spécificité. Il n'y avait notamment pas de différence pour le nombre de nouvelles lésions de taille seuil 3 mm comme référence en FUSION (AUC de 0,668, sensibilité de 30 % et spécificité de 96 %). Le seuil optimal était d'au moins 5 nouvelles lésions PIXYL  $\geq 3$  mm (AUC de 0,628, sensibilité de 39 % et spécificité de 82 %). PIXYL fournissait de nombreux autres marqueurs radiologiques, mais même l'utilisation de ratio des volumes par rapport au nombre de lésions PIXYL et la pondération des lésions ne permettait pas d'obtenir de meilleures performances de classification que FUSION. En effet, tout gain significatif de sensibilité était contrebalancé par une perte de spécificité (< 90 %) jugée incompatible avec un usage clinique en raison des conséquences potentielles de l'escalade thérapeutique exposant à plus de risques et étant limitée en nombre d'alternatives (45). C'était l'abaissement du seuil de nouvelles lésions FUSION à  $n \geq 2$  qui tendait aux meilleures performances (sensibilité de 42 % et spécificité de 92 %). Les méthodes de segmentation automatique dont PIXYL n'avaient jamais fait l'objet d'évaluations cliniques de leurs performances à identifier les patients en échec thérapeutique ou à risque de l'être. Il n'était donc possible de se référer qu'aux études de leur performance à obtenir une segmentation proche de celle de référence obtenue manuellement. Cependant, la définition de la segmentation de référence est une difficulté car il existe une large variabilité entre plusieurs segmentations manuelles réalisées par des experts (6). Une étude d' Egger et al. retrouvait une différence médiane de huit lésions soit un taux de faux positifs de 32,4 % pour une sensibilité de 67,6 % entre trois experts ayant segmenté la charge lésionnelle de 50 IRM (10). Notre critère de jugement clinique permettait de dépasser cette difficulté à définir une segmentation de référence avec une approche basée sur l'intérêt d'un usage en pratique clinique, c'est-à-dire à la manière des études ayant validé la pertinence de l'activité radiologique mesurée par le radiologue (46,47). De plus, les performances des méthodes de segmentation automatique sont difficilement comparables en raison de l'utilisation d'imageries IRM différentes dans chaque étude. Néanmoins, la variabilité de la segmentation entre différents experts et les moindres performances des méthodes automatiques sont constantes (6,10,22). Les méthodes de

segmentation supervisées sont dépendantes des imageries utilisées pour l'entraînement de la méthode. Il est important que les méthodes de segmentation automatiques soient robustes et reproductibles en dépit des difficultés à définir une segmentation de référence. Ainsi l'étude de trois méthodes automatiques, différentes de PIXYL, des IRM réalisées à deux reprises sur trois machines différemment paramétrées chez 10 patients, retrouvait un Dice-score (score de similarité) de  $0,71 \pm 0,18$  au mieux, avec une différence de volume segmenté de  $0,44 \pm 0,69$  ml et donc une reproductibilité encore imparfaite (22).

Le meilleur moyen de comparer notre étude avec la littérature était les analyses de concordance que l'on pouvait mettre en parallèle avec celles de similarité des segmentations. Dans notre étude, l'absence de corrélation forte entre le nombre de lésions ou l'activité  $\geq 3$  nouvelles lésions mesurés par la méthode FUSION et PIXYL, comme avec l'échec thérapeutique, était surprenante puisque les études de segmentation automatiques retrouvent en général une forte corrélation avec la segmentation de référence ( $> 0,9$ ) (10,22). PIXYL retrouvait plus de nouvelles lésions et leur nombre était mal corrélé avec celui de FUSION (Tau de Kendall à 0,26 pour la taille seuil de 3 mm). Les deux méthodes ne mesuraient donc pas tout à fait les mêmes lésions. Le challenge MICCAI 2016 a comparé différentes méthodes de segmentation automatiques, dont PIXYL, à la segmentation moyenne manuelle de 7 experts (6). Il reposait sur la segmentation des hypersignaux FLAIR de 53 IRM cérébrales de patients SEP acquises selon le protocole OFSEP sur 4 machines IRM de 3 centres hospitaliers différents. PIXYL se situait parmi les meilleures méthodes automatiques, qui avaient néanmoins toutes des performances moindres que les radiologues. Les coefficients de similarité (Dice-score et F-score) étaient de 48,9 % et 36 % pour respectivement les volumes et comptes lésionnels (vs 71,3 % [66,9 % - 78,2 %] et 78,7% [65,6 % - 89,3 %] pour les experts). Ces chiffres rendaient mieux compte du manque de concordance du compte lésionnel PIXYL avec la FUSION et l'échec thérapeutique dans notre étude. Ceci se traduisait aussi par la « distance de la surface », indicateur utilisé dans le challenge MICCAI 2016 pour refléter le volume segmenté en dehors de la segmentation de référence (distance parfaite de 0), qui était de 1,89 pour PIXYL par rapport à la moyenne de 0,66 [0,02 - 2,19] pour les experts.

Si l'on reprenait nos tableaux de concordance PIXYL / FUSION, la sensibilité de PIXYL pour détecter les lésions FUSION était de 90 % (27 / 30) et la spécificité de 25 % (13 / 53) pour le seuil d'une nouvelle lésion de taille seuil  $\geq 3$  mm. Elles étaient respectivement de 83 %

(10 / 12) et 61 % (43 / 71) pour le seuil  $\geq 3$  nouvelles lésions  $\geq 3$  mm. La relecture des masques de segmentation ne permettait pas d'affirmer que la sensibilité de PIXYL était parfaite puisque neuf patients (10,8 %) comportaient des nouvelles lésions détectées par la FUSION et pas par PIXYL mais il semblerait que la sensibilité de PIXYL se soit améliorée par rapport à la version utilisée lors du MICCAI 2016.

En effet dans ce challenge, la sensibilité de PIXYL par rapport au volume lésionnel (57,0 % versus 77,5 % [61,4 % - 87,4 %] pour les sept experts) comme pour le nombre de lésions (43,9 % versus 82,9 % [72,8 % - 89,9 %] des experts) était un point faible de toutes les méthodes automatiques (6). Elle semblait moindre que nos résultats. De plus dans notre étude, la sensibilité clinique des marqueurs PIXYL pour identifier l'échec thérapeutique était plutôt meilleure que FUSION. Les études concernant les méthodes de segmentation automatique se concentrent sur la segmentation de la charge lésionnelle globale à un temps défini. Or il existe un paradoxe radio-clinique avec une discordance entre la sévérité clinique et la charge lésionnelle. Ceci est notamment expliqué par la localisation stratégique des lésions et les atteintes mal identifiées en IRM conventionnelle de la substance blanche dite d'apparence normale (36). Pour le suivi des patients, l'évolution de la charge lésionnelle est un marqueur plus adapté pour juger de l'évolutivité de la maladie et de la réponse aux traitements. Une autre solution commerciale similaire à PIXYL, nommée LesionQuant, permet d'obtenir ces données d'évolutivité radiologique. Malheureusement, la seule étude d'évolutivité disponible pour cette méthode de segmentation automatique n'allait pas assez loin dans l'analyse de l'évolutivité clinique et radiologique en comparant les IRM à 1 et 5 ans du diagnostic (3). Parmi les 56 patients, 32 (57 %) avaient au moins une nouvelle lésion détectée par le radiologue, 17 de ces patients étaient également détectés par LesionQuant soit une sensibilité de 53 % et une spécificité de 75 % (6 patients stables détectés évolutifs par LesionQuant) (3). Même si la sensibilité de PIXYL semblait plus élevée que celle de LesionQuant, elle n'en différait pas ( $p = 0,25$ ) alors que sa spécificité était significativement plus faible pour le seuil d'une nouvelle lésion ( $p = 0,02$ ). Un nouveau challenge MICCAI a été organisé le 23 septembre 2021 en se focalisant sur ce point clé de la segmentation automatique qu'est l'évolutivité radiologique entre deux IRM chez 100 patients mais les résultats n'étaient pas encore disponibles (5). La version de PIXYL de notre étude semblait donc avoir gagné en sensibilité au détriment de la spécificité.

Dans le challenge MICCAI 2016, PIXYL avait une spécificité étonnamment élevée comme les autres méthodes testées ou les experts, mais cette spécificité calculée pour les volumes

segmentés rendait en réalité plutôt compte de l'efficacité à ne pas détecter trop de volume lésionnel non pertinent (6). Les lésions de la substance blanche ne concernant qu'une faible proportion du volume cérébral total, ce chiffre était peu informatif (99,9 %). L'indicateur le plus pertinent par rapport au nombre de lésions détectées était plutôt la précision (VPP) de 40,8 % pour PIXYL face aux 78,3 % [min 58 % - max 89,2 %] des sept experts. Notre précision de PIXYL, au seuil de trois nouvelles lésions  $\geq 3$  mm, pour identifier les patients répondant au même seuil FUSION semblait néanmoins plus faible avec 26,3 % (10 / 38). Notre étude ne retrouvait pas non plus une spécificité clinique élevée de PIXYL par rapport à FUSION pour identifier les patients avec échec thérapeutique (entre 50 % et 80 % selon le marqueur radiologique PIXYL au seuil considéré optimal par rapport au chiffre de 96 % pour la FUSION). Nos résultats étaient confortés par la relecture des masques de segmentation retrouvant beaucoup trop de faux positifs détectés (artefacts chez 63 patients (75,9 %) dont 31 avaient une IRM de FUSION interprétée stable). Ainsi, nos résultats étaient-ils en accord avec un probable manque de spécificité et de précision de PIXYL en pratique clinique à la différence des études de segmentation passées.

La spécificité clinique des nouvelles lésions PIXYL sans taille seuil était plus faible (54 % ;  $p < 0,001$ ) que celle de FUSION. L'augmentation de la spécificité du processus pathologique inflammatoire avec la taille des lésions était en accord avec la littérature et le seuil préconisé de 3 mm (21,52). En effet, les lésions aspécifiques, notamment d'origine vasculaire, peuvent être nombreuses et sont parfois difficiles à différencier des lésions inflammatoires démyélinisantes de la SEP (21). Les performances de discrimination des méthodes automatiques augmentent avec la taille des lésions (6) de même que leur spécificité (21). On pourrait donc penser que PIXYL détecterait plus facilement les lésions inflammatoires surtout avec application d'une taille seuil. Cependant notre étude n'a pas permis de retrouver une forte spécificité clinique de l'activité radiologique PIXYL sur la survenue d'un échec thérapeutique. L'expertise du radiologue assisté par la FUSION semblait apporter une meilleure spécificité pour détecter des lésions associées à la survenue d'un échec thérapeutique. Or, la spécificité des lésions détectées est importante pour un usage clinique de suivi de l'activité radiologique à risque d'activité clinique afin d'adapter précocement le traitement à l'activité de la maladie.

Le nombre de lésions PIXYL ne différait pas selon la prise d'un traitement contrairement au nombre de nouvelles lésions FUSION. Pourtant, le compte de lésions PIXYL avait une variabilité plus étendue et aurait dû permettre de mesurer plus finement l'activité radiologique en rapport avec la prise d'un traitement. Cette dispersion du nombre de lésions PIXYL associée à leur manque de spécificité étaient des explications probables. Un décompte exact des lésions est important dans l'évaluation de l'efficacité des traitements. L'absence de toute nouvelle lésion IRM est utilisée dans le score NEDA (No Evidence of Disease Activity) qui définit un contrôle complet de l'évolution de la SEP (39). Le NEDA est déjà difficile à maintenir dans la durée avec les évaluations radiologiques usuelles (0 % – 10 % à 10 ans pour la plupart des traitements) (19,43). Le NEDA serait encore plus difficile à atteindre lors de l'utilisation de méthodes automatiques comme PIXYL détectant plus de lésions. C'est une autre raison pour laquelle la spécificité des lésions détectées est indispensable pour un usage clinique et la présence de faux positifs en grand nombre est un problème à résoudre dans le développement de PIXYL.

Une partie des faux positifs était attribuée à des artefacts de contraste lors de la relecture des masques. Ces modifications de contraste proviennent souvent de paramètres d'acquisition d'IRM différents. Mais elles pourraient indiquer une activité à bas bruit de la maladie. Des lésions structurelles et le processus inflammatoire de la SEP s'étendent à des régions de substance blanche dite d'apparence normale en IRM conventionnelle. De discrètes modifications du contraste et de la texture de ces régions pourraient traduire une évolutivité radiologique (31). Des méthodes automatiques permettant l'analyse des modifications de texture de la substance blanche pourraient donc avoir un intérêt (55). Néanmoins, la pertinence clinique de cette activité radiologique n'est pas connue. L'utilisation d'IRM à très haut champ ( $> 3$  T) ou encore de paramètres d'acquisition de haute résolution (voxels  $< 1$  mm<sup>3</sup>) serait une autre piste d'amélioration des performances de segmentation automatiques avec un meilleur contraste signal / bruit. Néanmoins le volume de données à traiter dans ce cas semble encore trop important pour un usage clinique (temps et ressources informatiques nécessaires) et a nécessité une réduction de la résolution pour pouvoir être utilisé dans certaines études (22). L'utilisation de paramètres d'acquisition et de séquences spécialement développés pour les méthodes de segmentation automatique ou encore la combinaison de plusieurs séquences pourrait aussi améliorer leur performance (11,28).

Les seuils lésionnels optimaux détectés par une méthode de segmentation automatique comme PIXYL pouvaient différer de ceux utilisés en pratique radiologique (score de Sormani avec un seuil  $\geq 3$  nouvelles lésions  $\geq 3$  mm (47)), en raison de performances de segmentation différentes comme l'a montré notre étude avec des seuils optimaux différents selon la taille des lésions. La recherche de seuils lésionnels PIXYL optimaux pour les nombres et volumes des lésions donnait des performances de classification de l'échec thérapeutique non différentes de la référence (3 nouvelles lésions FUSION). Le seuil de la dimension  $y \geq 6$  mm ressortait pour les nouvelles lésions et pouvait s'appliquer à tous les axes pour des performances comparables à FUSION avec un seuil optimal d'une seule nouvelle lésion. Cette taille de 6 mm était plus importante que les 3 mm utilisés en pratique et validés dans la littérature (21). Nous avons fait l'hypothèse qu'une taille seuil plus élevée pour PIXYL aurait pu aider à compenser son manque de spécificité clinique. Néanmoins les performances pour une taille seuil des nouvelles lésions PIXYL de 3 mm (sensibilité de 39 % et spécificité de 82 %) et 6 mm (sensibilité de 42 % et spécificité de 96 %) étaient statistiquement proches même si la spécificité semblait augmenter. Les seuils optimaux considérés étaient d'ailleurs différents respectivement de  $n \geq 5$  et 1 lésions. Ce résultat suggérait que la taille des nouvelles lésions pourrait être intéressante pour identifier un risque d'échec thérapeutique. Une seule grosse lésion pourrait s'associer à un risque similaire à plusieurs petites lésions. Le caractère péjoratif des lésions volumineuses a été montré pour les lésions pseudo-tumorales de SEP (Tumefactive Demyelinating Lesion ou TDL) d'une taille supérieure à 50 mm et même supérieure ou égale à 35 mm (34,40). Les lésions les plus volumineuses sont également plus souvent associées à de l'œdème. En général, une manifestation pseudo-tumorale de SEP n'est pas associée à un risque d'échec thérapeutique plus important, le temps médian avant une seconde poussée étant de 5 ans dans une étude et le handicap à long terme similaire à ceux des autres patients SEP (34). Ces patients peuvent néanmoins être l'objet d'un traitement initial plus agressif selon une stratégie d'induction thérapeutique. La taille d'intérêt des lésions pourrait aussi différer selon leur localisation mais cela n'a pas pu être étudié de manière pertinente sur le plan statistique en raison de l'effectif de notre étude.

La combinaison de plusieurs marqueurs PIXYL permettait d'obtenir un modèle de régression logistique intéressant (AUC de 0,758, sensibilité de 55 % et spécificité de 90 % au seuil optimal de 0,499). Ses performances ne différaient pas significativement du seuil de référence de 3 lésions FUSION. Le nombre de nouvelles lésions infratentorielles était le

marqueur PIXYL le plus robuste qui restait statistiquement significatif en analyse multivariée. La localisation des lésions est importante pour le retentissement clinique. La localisation infratentorielle est un facteur péjoratif connu pour toutes les formes de SEP, le risque de convertir un CIS en SEP et d'évoluer vers une forme progressive (14,26,29). Néanmoins, il faut rester prudent par rapport aux performances de classification de PIXYL. En effet, les autres marqueurs radiologiques manquaient de significativité forte par rapport au nombre de nouvelles lésions infratentorielles. Il existait même un non-sens avec un nombre de lésions élargies associé à un odds ratio  $< 1$  dans les modèles de régression logistique et qui pouvait même être statistiquement significatif. Ceci était totalement discordant avec la littérature qui accorde la même valeur à toutes les lésions évolutives, qu'elles soient nouvelles ou franchement élargies (45).

PIXYL ne pouvait pas se substituer à l'expertise d'un radiologue mais pourrait être complémentaire du fait du rapport sensibilité / spécificité plutôt inversé par rapport à FUSION. La combinaison des meilleurs marqueurs radiologiques PIXYL et FUSION dans un modèle de régression logistique global obtenait de meilleures performances que le nombre de lésions FUSION seul (AUC de 0,836). La sensibilité était de 64 % et la spécificité de 96 % au seuil optimal de probabilité d'un échec fixé à 0,548. La valeur clinique ajoutée des méthodes de segmentation automatique pour identifier l'échec thérapeutique n'avait jamais été étudiée auparavant et n'a donc pas pu être comparée à nos résultats.

La répartition différente des marqueurs radiologiques FUSION selon l'échec thérapeutique était attendue puisque l'activité radiologique est connue comme marqueur précoce de l'activité de la maladie (46). Ceci est surtout valable pour le seuil d'au moins 3 nouvelles lésions qui est reconnu dans la littérature pour les traitements par interférons (47) et était appliqué par le clinicien pour motiver la décision d'un changement de traitement chez 11 patients (33 % des échecs). De façon surprenante, les distributions des nombres de lésions PIXYL semblaient moins différer et les seuils habituels d'une ou trois nouvelles lésions étaient moins significatifs qu'en FUSION. C'était notamment le cas pour les nombres de nouvelles lésions PIXYL (même avec une taille seuil de 3 mm) qui ne différaient pas selon la survenue d'un échec thérapeutique global. Les différences entre les groupes avec et sans échec étaient bien plus marquées pour les marqueurs radiologiques FUSION et PIXYL en cas d'échec entre les deux IRM. Par contre, elles n'étaient pas significatives en cas d'échec survenu dans l'année suivant l'IRM de

FUSION. Ceci était plutôt logique puisque les différences de segmentation entre deux IRM mesurent ce qui s'est passé entre les deux imageries. Cette activité peut être prédictive de la survenue future d'un échec (46,47). Néanmoins, cette prédiction est moins forte. Notre étude n'avait probablement pas la puissance statistique nécessaire compte tenu du nombre réduit d'échecs survenus après l'IRM de FUSION ( $n = 16$  soit 19,3 % des patients soit un peu moins de la moitié des 33 patients avec échec observés dans l'étude). De plus, l'activité inflammatoire de la SEP varie dans le temps (8). En effet, les adaptations thérapeutiques chez les patients les plus actifs entre deux IRM renforcent la pression anti-inflammatoire avec l'escalade thérapeutique.

La vitesse de survenue d'un échec thérapeutique dans l'année suivant l'IRM de FUSION montrait encore l'intérêt du seuil de 2 nouvelles lésions FUSION pour améliorer la sensibilité sans perdre trop en spécificité. Le test du logrank montrait une relation forte avec la vitesse de survenue d'un échec et ce seuil alors que ce n'était pas le cas pour le seuil de référence de 3 lésions FUSION. La relation n'était pas significative pour les seuils des comptes de lésions PIXYL. Seul le volume élargi ressortait dans les marqueurs non composites PIXYL pour distinguer les patients présentant un échec plus rapidement dans l'année suivant l'IRM de FUSION. Les modèles de régression *PIXYL* et *GLOBAL* étaient confortés par les courbes de survie montrant une association forte avec la rapidité de survenue d'un échec au seuil optimal sélectionné auparavant. La vitesse de survenue de l'échec est un indicateur important qui a été utilisé pour valider les scores de Rio ou Sormani (47).

La prudence s'imposait dans l'interprétation des données PIXYL. Elles ne sont pas encore bien validées sur le plan radiologique comme clinique. Notamment, après les analyses de concordance, laquelle était faible entre le nombre de nouvelles lésions PIXYL et FUSION. La concordance était également faible entre le nombre de nouvelles lésions PIXYL comme FUSION et l'échec thérapeutique. Mais cette concordance mesurée par le Kappa de Cohen dépendait beaucoup des sensibilités et spécificités des marqueurs étudiés. Les ratio sensibilité / spécificité n'étaient pas parfaits et étaient inversés entre PIXYL et FUSION aboutissant à ces concordances similairement faibles. La relecture des masques de segmentation a encore suscité plus de prudence quant à la pertinence des lésions détectées par PIXYL. En effet, il existait beaucoup de faux positifs (63 patients soit 75,9 % des IRM de l'étude)

principalement causés par des lésions artéfactuelles dont la non-prise en compte ne semblait pas facile.

Sur le plan méthodologique, il s'agissait d'une étude rétrospective bien que les données cliniques aient pu être recueillies prospectivement grâce à la base de données EDMUS. Le caractère monocentrique a limité le nombre de patients disponibles ayant bénéficié de la procédure de FUSION. La puissance statistique a donc pu manquer pour mettre en évidence une différence de performance de PIXYL comme le montrait la largeur des intervalles de confiance des AUC (un peu plus de 0,2). Cependant même une petite différence pourrait avoir un intérêt clinique compte tenu du manque de sensibilité de l'activité radiologique mesurée en FUSION pour identifier les patients subissant un échec thérapeutique.

Les IRM reflétaient la pratique en vie réelle des paramètres d'acquisition utilisés qui pouvaient varier bien qu'adaptés des recommandations de l'OFSEP. Cette variabilité des paramètres d'acquisition notamment reflétée par les 36 patients (43,4 %) ayant eu une discordance entre un FLAIR, avec ou sans saturation de graisse, a pu perturber le fonctionnement de PIXYL dans la gestion des différences de contraste (bien qu'un pré-traitement avec normalisation de l'image fasse partie de la procédure de segmentation PIXYL). L'utilisation de méthodes automatiques comme PIXYL est en effet limitée par la nécessité d'imageries IRM de bonne qualité avec des séquences 3D. Les séquences 3DT1 et 3DFLAIR font partie du protocole d'acquisition recommandé par l'OFSEP pour les IRM cérébrales de suivi des patients SEP (2). Dans notre étude, seuls 21 / 83 (25,3 %) des patients sélectionnés pour une IRM de FUSION (qui nécessite un 3DFLAIR) avaient eu un 3DT1 lors des 2 IRM de suivi. Ces séquences n'étaient donc pas toujours réalisées en pratique à l'époque de l'étude en partie à cause d'un temps d'acquisition plus long que les séquences 2D.

La durée de suivi disponible en ayant inclus les IRM cérébrales réalisées avec la FUSION la plus récente ne permettait pas d'explorer la survenue d'un échec thérapeutique au-delà d'un an en raison du nombre de patients pour lesquels la donnée n'était pas disponible par rapport à l'effectif inclus (suivi trop précoce et perdus de vue). L'intervalle de réalisation des deux IRM cérébrales était variable bien qu'il soit proche de la durée habituelle d'un an reflétant les aléas de la pratique courante à la différence des protocoles de recherche.

Un biais de mesure de l'échec thérapeutique est possible pour les 21 patients (25,3 %) pour lesquels l'IRM de FUSION correspondait à une rebaseline. En effet, le nouveau traitement peut

arriver à efficacité sur le plan clinique, donc sans poussée ni aggravation de l'EDSS alors qu'une activité radiologique précoce a pu persister dans cet intervalle de temps. Notamment, car l'activité radiologique est plus importante que l'activité clinique mesurable avec la notion de 10 lésions IRM asymptomatiques pour une poussée clinique (9). D'où la nécessité d'IRM de rebaseline comme nouvelle référence pour les IRM ultérieures, l'évolution radiologique à ce moment étant en général tolérée. Ainsi, une IRM de rebaseline ne sert pas à évaluer l'efficacité d'un traitement. L'activité radiologique d'une rebaseline avec apparition d'au moins trois nouvelles lésions ne sera pas considérée comme un échec thérapeutique motivant un changement de traitement pour inefficacité. Il était néanmoins difficile d'exclure ce nombre significatif de patients concernés pour garder une population étudiée suffisante. De plus le biais de mesure de l'échec thérapeutique ne pouvait au plus concerner que les 9 patients (10,8 %) pour lesquels aucun échec thérapeutique n'est survenu dans ce contexte de rebaseline.

L'utilisation d'un critère composite pour définir l'échec thérapeutique permettait d'avoir un maximum d'événements mais était critiquable en raison de l'inclusion de l'arrêt de traitement pour inefficacité. En effet, cette dernière repose sur l'activité clinique de la maladie mais également sur l'activité radiologique qui pouvait provenir de l'IRM de FUSION ou de l'IRM comparée, les patients ne pouvant qu'exceptionnellement avoir eu une autre IRM dans la période de suivi limitée à un an de l'IRM de FUSION. Néanmoins, cette définition composite de l'échec thérapeutique a déjà été utilisée dans de précédentes études (47) et la FUSION étant une méthode d'interprétation IRM de référence validée en pratique, PIXYL devait au moins pouvoir détecter ces patients pour lesquels la FUSION avait modifié le traitement.

La portée clinique d'identification radiologique de l'échec thérapeutique est surtout intéressante pour prédire un échec à venir. Malheureusement, le nombre d'échecs thérapeutiques survenus après l'IRM de FUSION était moindre ( $n = 16$ ). L'analyse de sous-groupe sur l'échec ultérieur à l'IRM de FUSION ne retrouvait pas de différence significative des marqueurs radiologiques PIXYL comme FUSION par rapport aux patients sans échec et les performances pour identifier les patients à risque d'échec futur semblaient moindres que pour identifier l'échec global ou survenu entre les deux IRM. Le nombre d'événements était probablement trop limité pour bien appréhender les performances de prédiction de l'échec ultérieur à l'IRM de FUSION. Augmenter la période de suivi aurait exposé au facteur confondant d'une modification de traitement sans inefficacité.

## CONCLUSION ET PERSPECTIVES

Cette étude était la première évaluation clinique d'une méthode de segmentation automatique pour comparaison d'IRM cérébrales dans le suivi de patients atteints de SEP RR. Les performances des marqueurs PIXYL synthétisées par l'AUC sensibilité / spécificité n'étaient pas supérieures à FUSION pour identifier les patients avec échec thérapeutique dans l'année suivant l'IRM de FUSION. Tout gain de sensibilité notable se faisait au prix d'une perte de spécificité incompatible avec un usage clinique. Néanmoins, la combinaison de FUSION avec PIXYL a permis la construction d'un modèle de régression logistique dont les performances pouvaient améliorer celles de la FUSION seule. Par rapport au seuil de référence de 'trois nouvelles lésions FUSION', le *Modèle GLOBAL* obtenu améliorait l'AUC de 25 % (de 0,668 à 0,836) et doublait la sensibilité de 30 % à 64 % pour une spécificité inchangée de 96 %. Ces performances permettaient d'identifier 33 % de patients supplémentaires avec échec thérapeutique.

Cependant, la prudence s'applique aux conclusions de cette étude exploratoire. En effet, les analyses de concordance entre les marqueurs PIXYL et FUSION et l'échec thérapeutique retrouvaient de mauvaises corrélations entre tous ces marqueurs. De plus, la relecture des masques de segmentation PIXYL retrouvait beaucoup de faux positifs dont des artefacts de contraste de signification douteuse.

PIXYL est encore en développement comme de nombreuses autres méthodes de segmentation automatique d'IRM. Aucune n'est pour l'instant validée pour l'utilisation en pratique quotidienne ni même utilisée largement. Leur but est de simplifier l'analyse des IRM de suivi des patients SEP. Elles devraient donc apporter un gain de temps, de performance, de reproductibilité et d'objectivité. Dans l'état actuel de la version étudiée, la relecture des masques de segmentation automatique majore le temps d'interprétation de l'imagerie si la méthode détecte trop de faux positifs.

Notre étude exploratoire suggérait que PIXYL pouvait encore être amélioré. Ses performances de segmentation devraient être précisées pour bien définir les limites de la méthode.

Il est possible que le manque de sensibilité de la FUSION pour identifier l'échec thérapeutique provienne des limites du nombre de lésions évolutives comme marqueur

radiologique. Les méthodes automatiques devraient donc également explorer la détection de modifications radiologiques mal analysées par l'œil humain. Par exemple, l'aspect de la substance blanche d'apparence normale ou la volumétrie / atrophie pourraient compléter la lecture radiologique. Les méthodes automatiques ont été principalement développées pour segmenter alors que l'application clinique est la comparaison pour recherche d'évolutivité. Il serait intéressant d'élaborer des méthodes directement construites sur l'analyse différentielle et l'identification des paramètres cliniques associés, notamment par deep learning (apprentissage profond par intelligence artificielle). Les paramètres d'acquisition d'IRM pourraient aussi être optimisés pour les méthodes de segmentation automatique même si cela réduirait l'usage à des imageries acquises dans ce but à moins que ces paramètres soient l'objet de nouvelles recommandations pour que toute imagerie de suivi puisse bénéficier des méthodes de segmentation automatique.

## **BIBLIOGRAPHIE**

1. Ashton EA, Takahashi C, Berg MJ, Goodman A, Totterman S, Ekholm S. Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI. *J Magn Reson Imaging*. mars 2003;17(3):300-8.
2. Brisset J-C, Kremer S, Hannoun S, Bonneville F, Durand-Dubief F, Tourdias T, et al. New OFSEP recommendations for MRI assessment of multiple sclerosis patients: Special consideration for gadolinium deposition and frequent acquisitions. *J Neuroradiol*. juin 2020;47(4):250-8.
3. Brune S, Høgestøl EA, Cengija V, Berg-Hansen P, Sowa P, Nygaard GO, et al. LesionQuant for Assessment of MRI in Multiple Sclerosis-A Promising Supplement to the Visual Scan Inspection. *Front Neurol*. 2020;11:546744.
4. Canty A, Ripley B. boot: Bootstrap R (S-Plus) Functions [Internet]. 2021. Disponible sur: <https://CRAN.R-project.org/package=boot>
5. Cervenansky F, Commowick O, Cotton F, Dojat M, Edan G. Multiple sclerosis new lesions segmentation challenge. Zenodo; 3 mars 2021 [cité 1 sept 2021]; Disponible sur: <https://zenodo.org/record/4575409>
6. Commowick O, Istace A, Kain M, Laurent B, Leray F, Simon M, et al. Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure. *Sci Rep*. 12 sept 2018;8(1):13650.
7. Compston A, Coles A. Multiple sclerosis. *Lancet*. 25 oct 2008;372(9648):1502-17.
8. Confavreux C, Vukusic S. The clinical course of multiple sclerosis. *Handb Clin Neurol*. 2014;122:343-69.
9. Dobson R, Giovannoni G. Multiple sclerosis - a review. *Eur. J. Neurol*. 2019;26(1):27-40.
10. Egger C, Opfer R, Wang C, Kepp T, Sormani MP, Spies L, et al. MRI FLAIR lesion segmentation in multiple sclerosis: Does automated segmentation hold up with manual annotation? *Neuroimage Clin*. 2017;13:264-70.
11. Essa E, Aldesouky D, Hussein SE, Rashad MZ. Neuro-fuzzy patch-wise R-CNN for multiple sclerosis segmentation. *Med Biol Eng Comput*. sept 2020;58(9):2161-75.
12. Falissard B. psy: Various procedures used in psychometry [Internet]. 2012. Disponible sur: <https://CRAN.R-project.org/package=psy>
13. Filippi M, Preziosa P, Banwell BL, Barkhof F, Ciccarelli O, De Stefano N, et al. Assessment of lesions on magnetic resonance imaging in multiple sclerosis: practical guidelines. *Brain*. 1 juill 2019;142(7):1858-75.
14. Filippi M, Preziosa P, Barkhof F, Chard DT, De Stefano N, Fox RJ, et al. Diagnosis of Progressive Multiple Sclerosis From the Imaging Perspective: A Review. *JAMA Neurol*. 1 mars 2021;78(3):351-64.

15. Forbes F, Doyle S, Garcia-Lorenzo D, Barillot C, Dojat M. A Weighted Multi-Sequence Markov Model For Brain Lesion Segmentation. Proceedings of the [Internet]. Chia Laguna Resort, Sardinia, Italy: Neil Lawrence; 2010. p. 225-32. Disponible sur: <http://proceedings.mlr.press/v9/forbes10a/forbes10a.pdf>
16. Foulon S, Maura G, Dalichampt M, Alla F, Debouverie M, Moreau T, et al. Prevalence and mortality of patients with multiple sclerosis in France in 2012: a study based on French health insurance data. *J. Neurol.* juin 2017;264(6):1185-92.
17. Galletto Pregliasco A, Collin A, Guéguen A, Metten MA, Aboab J, Deschamps R, et al. Improved Detection of New MS Lesions during Follow-Up Using an Automated MR Coregistration-Fusion Method. *AJNR Am J Neuroradiol.* juill 2018;39(7):1226-32.
18. García-Lorenzo D, Francis S, Narayanan S, Arnold DL, Collins DL. Review of automatic segmentation methods of multiple sclerosis white matter lesions on conventional magnetic resonance imaging. *Med Image Anal.* janv 2013;17(1):1-18.
19. Gasperini C, Prosperini L, Tintoré M, Sormani MP, Filippi M, Rio J, et al. Unraveling treatment response in multiple sclerosis: A clinical and MRI challenge. *Neurology.* 22 janv 2019;92(4):180-92.
20. Giovannoni G, Butzkueven H, Dhib-Jalbut S, Hobart J, Kobelt G, Pepper G, et al. Brain health: time matters in multiple sclerosis. *Mult Scler Relat Disord.* sept 2016;9 Suppl 1:S5-48.
21. Grahl S, Pongratz V, Schmidt P, Engl C, Bussas M, Radetz A, et al. Evidence for a white matter lesion size threshold to support the diagnosis of relapsing remitting multiple sclerosis. *Mult Scler Relat Disord.* avr 2019;29:124-9.
22. Jain S, Sima DM, Ribbens A, Cambron M, Maertens A, Van Hecke W, et al. Automatic segmentation and volumetry of multiple sclerosis brain lesions from MR images. *Neuroimage Clin.* 2015;8:367-75.
23. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using « ggplot2 » [Internet]. 2020. Disponible sur: <https://CRAN.R-project.org/package=survminer>
24. Kaunzner UW, Gauthier SA. MRI in the assessment and monitoring of multiple sclerosis: an update on best practice. *Ther Adv Neurol Disord.* juin 2017;10(6):247-61.
25. Kobelt G, Thompson A, Berg J, Gannedahl M, Eriksson J, MSCOI Study Group, et al. New insights into the burden and costs of multiple sclerosis in Europe. *Mult Scler.* juill 2017;23(8):1123-36.
26. Kohler M, Kohler E, Vrech C, Pappolla A, Miguez J, Patrucco L, et al. Aggressive multiple sclerosis in Argentina: Data from the nationwide registry RelevaEM. *J Clin Neurosci.* juill 2021;89:360-4.
27. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology.* nov 1983;33(11):1444-52.

28. Le M, Tang LYW, Hernández-Torres E, Jarrett M, Brosch T, Metz L, et al. FLAIR2 improves LesionTOADS automatic segmentation of multiple sclerosis lesions in non-homogenized, multi-center, 2D clinical magnetic resonance images. *Neuroimage Clin.* 2019;23:101918.
29. Lebrun-Frenay C, Kantarci O, Siva A, Sormani MP, Pelletier D, Okuda DT, et al. Radiologically Isolated Syndrome: 10-Year Risk Estimate of a Clinical Event. *Ann Neurol.* août 2020;88(2):407-17.
30. Lemon J, Grosjean P. prettyR: Pretty Descriptive Stats [Internet]. 2019. Disponible sur: <https://CRAN.R-project.org/package=prettyR>
31. Loizou CP, Pantzaris M, Pattichis CS. Normal appearing brain white matter changes in relapsing multiple sclerosis: Texture image and classification analysis in serial MRI scans. *Magn Reson Imaging.* nov 2020;73:192-202.
32. Lublin FD, Reingold SC. Defining the clinical course of multiple sclerosis: results of an international survey. National Multiple Sclerosis Society (USA) Advisory Committee on Clinical Trials of New Agents in Multiple Sclerosis. *Neurology.* avr 1996;46(4):907-11.
33. Lublin FD, Reingold SC, Cohen JA, Cutter GR, Sørensen PS, Thompson AJ, et al. Defining the clinical course of multiple sclerosis: the 2013 revisions. *Neurology.* 15 juill 2014;83(3):278-86.
34. Lucchinetti CF, Gavrilova RH, Metz I, Parisi JE, Scheithauer BW, Weigand S, et al. Clinical and radiographic spectrum of pathologically confirmed tumefactive multiple sclerosis. *Brain.* 1 juill 2008;131(7):1759-75.
35. Meyer D, Zeileis A, Hornik K. vcd: Visualizing Categorical Data. 2020.
36. Mollison D, Sellar R, Bastin M, Mollison D, Chandran S, Wardlaw J, et al. The clinico-radiological paradox of cognitive function and MRI burden of white matter lesions in people with multiple sclerosis: A systematic review and meta-analysis. *PLoS One.* 2017;12(5):e0177727.
37. Moraal B, Meier DS, Poppe PA, Geurts JJG, Vrenken H, Jonker WMA, et al. Subtraction MR images in a multiple sclerosis multicenter clinical trial setting. *Radiology.* févr 2009;250(2):506-14.
38. Pardo G, Jones DE. The sequence of disease-modifying therapies in relapsing multiple sclerosis: safety and immunologic considerations. *J. Neurol.* déc 2017;264(12):2351-74.
39. Parks NE, Flanagan EP, Lucchinetti CF, Wingerchuk DM. NEDA treatment target? No evident disease activity as an actionable outcome in practice. *J Neurol Sci.* 15 déc 2017;383:31-4.
40. Plowman RS, Varma H. Prognostic factors in Tumefactive demyelinating lesions: A retrospective study. *J Neurol Sci.* 15 sept 2021;428:117591.
41. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2020. Disponible sur: <https://www.R-project.org>

42. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez J-C, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*. 17 mars 2011;12(1):77.
43. Rotstein DL, Healy BC, Malik MT, Chitnis T, Weiner HL. Evaluation of no evidence of disease activity in a 7-year longitudinal multiple sclerosis cohort. *JAMA Neurol*. févr 2015;72(2):152-8.
44. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. Boston, MA: RStudio, PBC.; 2020. Disponible sur: <http://www.rstudio.com/>
45. Smith AL, Cohen JA, Hua LH. Therapeutic Targets for Multiple Sclerosis: Current Treatment Goals and Future Directions. *Neurotherapeutics*. oct 2017;14(4):952-60.
46. Sormani MP, Bruzzi P. MRI lesions as a surrogate for relapses in multiple sclerosis: a meta-analysis of randomised trials. *Lancet Neurol*. juill 2013;12(7):669-76.
47. Sormani MP, Gasperini C, Romeo M, Rio J, Calabrese M, Cocco E, et al. Assessing response to interferon- $\beta$  in a multicenter dataset of patients with MS. *Neurology*. 12 juill 2016;87(2):134-40.
48. Terry M. Therneau, Patricia M. Grambsch. *Modeling Survival Data: Extending the Cox Model*. New York: Springer; 2000.
49. Therneau TM. A Package for Survival Analysis in R [Internet]. 2020. Disponible sur: <https://CRAN.R-project.org/package=survival>
50. Thompson AJ, Banwell BL, Barkhof F, Carroll WM, Coetzee T, Comi G, et al. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*. févr 2018;17(2):162-73.
51. Venables WN, Ripley BD. *Modern Applied Statistics with S* [Internet]. Fourth. New York: Springer; 2002. Disponible sur: <http://www.stats.ox.ac.uk/pub/MASS4/>
52. Wattjes MP, Ciccarelli O, Reich DS, Banwell B, de Stefano N, Enzinger C, et al. 2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis. *Lancet Neurol*. août 2021;20(8):653-70.
53. Wickham H. *ggplot2: Elegant Graphics for Data Analysis* [Internet]. Springer-Verlag New York; 2016. Disponible sur: <https://ggplot2.tidyverse.org>
54. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, et al. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *NeuroImage*. 2006;31(3):1116-28.
55. Zhang Y. MRI texture analysis in multiple sclerosis. *Int J Biomed Imaging*. 2012;2012:762804.
56. PIXYL.MS [Internet]. Disponible sur: <https://pixyl.ai/multiple-sclerosis/>

## ANNEXES

Annexe 1. Sous-scores Fonctionnels de l'EDSS .....	85
Annexe 2. Courbes sensibilité / spécificité du nombre de nouvelles lésions PIXYL.....	86
Annexe 3. Courbes sensibilité / spécificité du nombre de lésions élargies PIXYL.....	86
Annexe 4. Courbes sensibilité / spécificité du nombre de lésions évolutives PIXYL.....	87
Annexe 5. Courbes sensibilité / spécificité du volume des lésions évolutives PIXYL .....	87
Annexe 6. Courbes sensibilité / spécificité du volume des lésions nouvelles PIXYL .....	88
Annexe 7. Courbes sensibilité / spécificité du volume des lésions élargies PIXYL .....	88
Annexe 8. Tableau des performances de différents marqueurs radiologiques PIXYL comparées à celles du nombre de lésion comptées en FUSION pour identifier la survenue d'un échec thérapeutique entre les deux IRM .....	89
Annexe 9. Tableau des performances de différents marqueurs radiologiques PIXYL comparées à celles du nombre de lésion comptées en FUSION pour identifier la survenue d'un échec thérapeutique après l'IRM de FUSION .....	90
Annexe 10. Seuils lésionnels PIXYL optimaux considérés par rapport au compte lésionnel pour identifier un échec entre les deux IRM .....	91
Annexe 11. Seuils lésionnels PIXYL optimaux considérés par rapport au volume lésionnel pour identifier un échec entre les deux IRM .....	91
Annexe 12. Seuils lésionnels PIXYL optimaux considérés par rapport au compte lésionnel pour identifier un échec postérieur à l'IRM de FUSION.....	92
Annexe 13. Seuils lésionnels PIXYL optimaux considérés par rapport au volume lésionnel pour identifier un échec postérieur à l'IRM de FUSION.....	92

<b>Fonction pyramidale</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Signes anormaux sans handicap
2	Handicap minime
3	Paraparésie ou hémiparésie légère ou modérée
4	Paraparésie ou hémiparésie marquée ; tétraparésie modérée ; ou monoplégie
5	Paraplégie, hémiplégie, ou tétraparésie marquée
6	Tétraplégie

<b>Fonction cérébelleuse</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Anormal sans handicap
2	Légère ataxie
3	Ataxie modérée des membres ou du tronc
4	Ataxie sévère des quatre membres
5	Impossibilité de réaliser des mouvements coordonnés en raison de l'ataxie
X	À ajouter quand le déficit moteur (grade 3 ou plus du score pyramidal) interfère avec l'examen

<b>Fonction sensitive</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Diminution du sens vibratoire ou seulement altération de la reconnaissance de figure dessinée sur la peau, affectant un ou deux membres
2	Légère diminution au touché-piqué ou au sens de position, et/ou diminution modérée du sens vibratoire, dans un ou deux membres ; ou diminution du sens vibratoire (ou de la reconnaissance de figure dessinée sur la peau) dans trois ou quatre membres
3	Diminution modérée au touché-piqué ou au sens de position, et/ou perte du sens vibratoire, dans un ou deux membres ; ou diminution légère du sens tactile nociceptif et/ou diminution modérée de tous les tests proprioceptifs de trois ou quatre membres
4	Diminution marquée au touché-piqué ou perte de la proprioception, isolées ou associées dans un ou deux membres ; ou diminution modérée au touché-piqué et/ou diminution sévère de la proprioception dans plus de deux membres
5	Perte (essentiellement) de la sensibilité dans un ou deux membres ; ou diminution modérée au touché-piqué et/ou perte de la proprioception de l'ensemble du corps en-dessous de la tête
6	Perte de la sensibilité en-dessous de la tête

<b>Fonction du tronc cérébral</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Signes uniquement

2	Nystagmus modéré ou autre handicap léger
3	Nystagmus sévère, faiblesse extra-oculaire marquée, ou handicap modéré d'autres nerfs crâniens
4	Dysarthrie marquée ou autre handicap marqué
5	Incapacité de déglutir ou parler

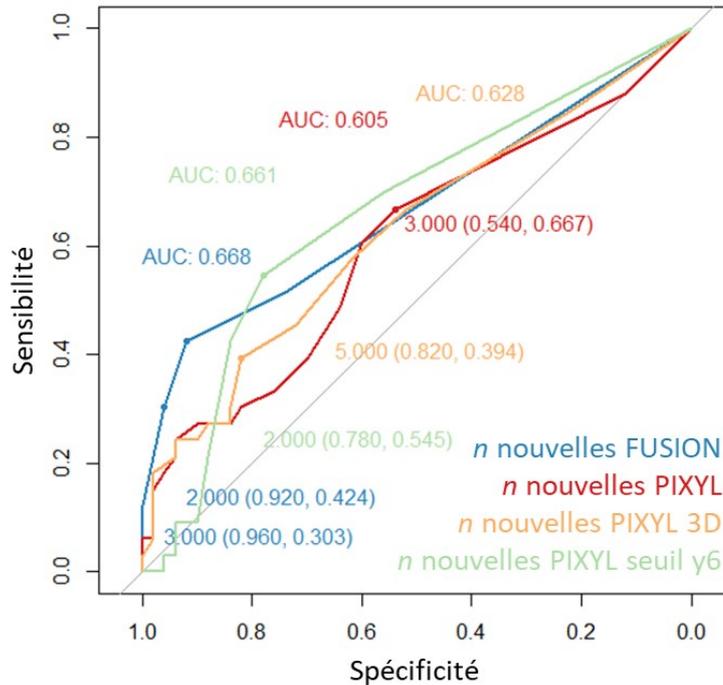
<b>Fonction sphinctérienne</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Légère dysurie, impériosité ou rétention
2	Dysurie, impériosité modérées, rétention urinaire ou constipation modérée, ou incontinence urinaire rare
3	Incontinence urinaire fréquente
4	Nécessite des sondages quasi-constants
5	Perte de fonction urinaire

<b>Fonction visuelle</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Scotome avec acuité visuelle corrigée > 0,67
2	Scotome sur l'œil le plus atteint avec une acuité visuelle entre 0,34 et 0,67
3	Large scotome sur l'œil le plus atteint, ou réduction du champ visuel, avec acuité visuelle comprise entre 0,2 et 0,33
4	Œil le plus atteint avec diminution marquée du champ visuel, acuité visuelle entre 0,1 et 0,2 ; ou grade 3 avec acuité visuelle du meilleur œil < 0,3
5	Acuité visuelle de l'œil le plus atteint < 0,1 ; ou grade 4 avec acuité visuelle du meilleur œil < 0,3
6	Grade 5 avec acuité visuelle du meilleur œil < 0,3
X	À ajouter quand il y a pâleur temporale

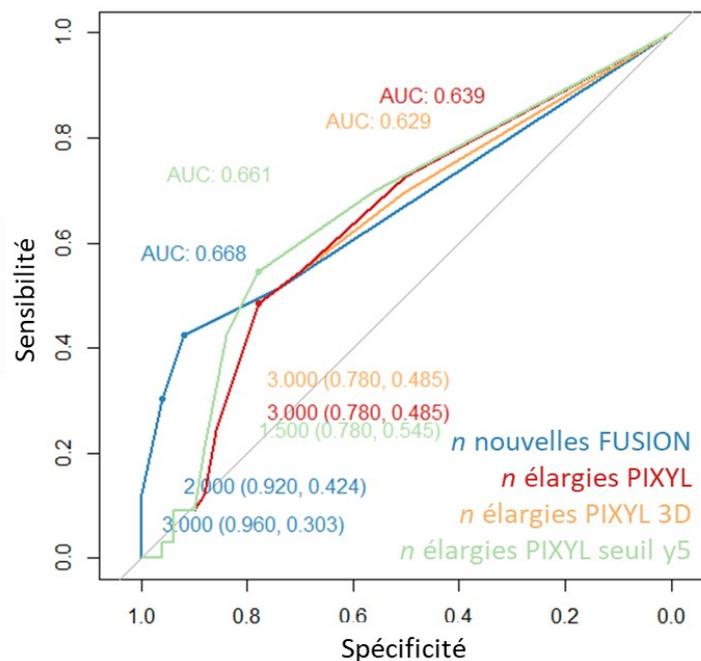
<b>Fonction cérébrale</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Altération isolée de l'humeur (n'affecte pas le score EDSS)
2	Diminution légère de l'idéation
3	Diminution modérée de l'idéation
4	Diminution marquée de l'idéation
5	Démence

<b>Autres fonctions</b>	
<u>Score</u>	<u>Critères</u>
0	Normal
1	Toute autre perturbation neurologique attribuable à la SEP

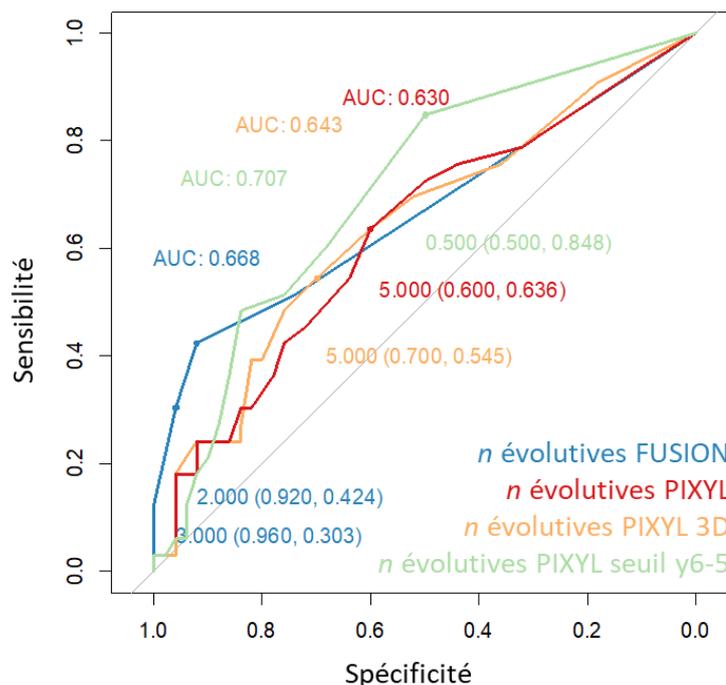
Annexe 1. Sous-scores Fonctionnels de l'EDSS adaptés de la traduction sur [www.edmus.org](http://www.edmus.org)  
© 2009 EDMUS Coordinating Center (27)



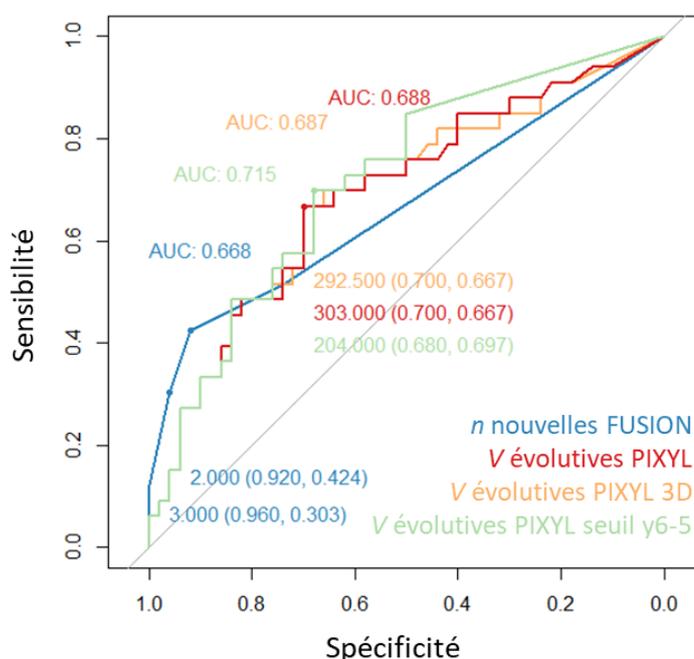
Annexe 2. Courbes sensibilité / spécificité du nombre de nouvelles lésions PIXYL comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité ;  $n$  : nombre de lésions)



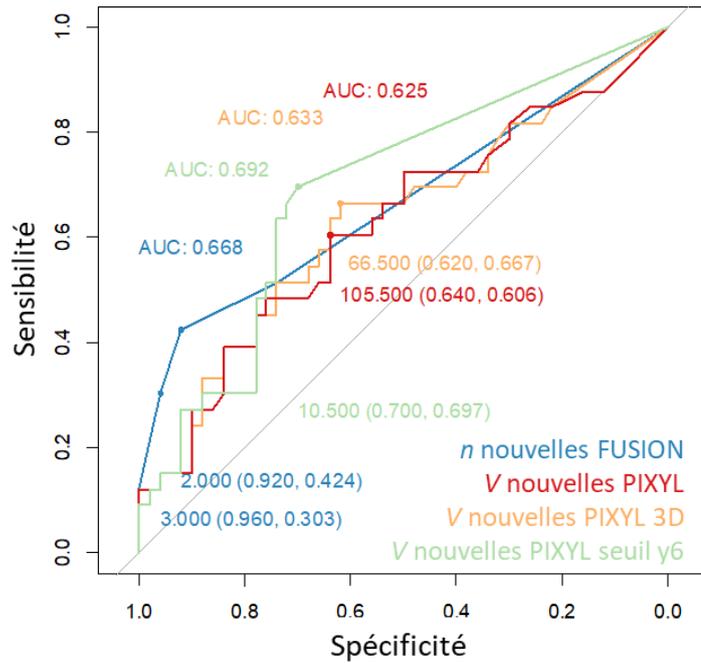
Annexe 3. Courbes sensibilité / spécificité du nombre de lésions élargies PIXYL comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité ;  $n$  : nombre de lésions)



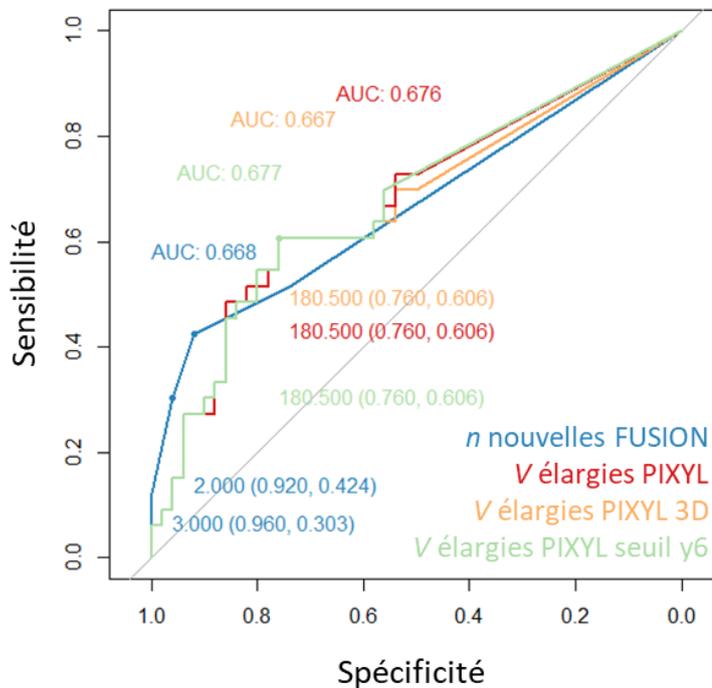
Annexe 4. Courbes sensibilité / spécificité du nombre de lésions évolutives PIXYL comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité ;  $n$  : nombre de lésions)



Annexe 5. Courbes sensibilité / spécificité du volume des lésions évolutives PIXYL comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité ;  $n$  : nombre de lésions ;  $V$  : Volume des lésions)



Annexe 6. Courbes sensibilité / spécificité du volume des lésions nouvelles PIXYL comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité ;  $n$  : nombre de lésions ;  $V$  : Volume des lésions)



Annexe 7. Courbes sensibilité / spécificité du volume des lésions élargies PIXYL comparées à la référence du nombre de lésions FUSION pour l'identification de l'échec thérapeutique global survenu à un an de l'IRM de FUSION (AUC : aire sous la courbe et seuil optimal pour le rapport sensibilité / spécificité rapporté sur la courbe avec coordonnées correspondant à la spécificité en fonction de la sensibilité ;  $n$  : nombre de lésions ;  $V$  : Volume des lésions)

Marqueur radiologique	AUC [IC95%]	p	Seuil	Se	Spé	VPP	VPN	Exact
Nouvelle lésion FUSION	0,717 [0,596-0,838]	1	3	0,4	0,966	0,833	0,789	0,795
Nouvelle lésion FUSION	0,717 [0,596-0,838]	1	2	<b>0,48</b>	0,897	0,667	<b>0,8</b>	0,771
Nouvelle lésion	0,649 [0,513-0,786]	0,411	11	0,36	0,914	0,643	0,768	0,747
Nouvelle lésion 3D*	0,689 [0,561-0,817]	0,718	5	<b>0,48</b>	<u>0,828</u>	0,545	0,787	0,723
Lésion élargie	0,714 [0,598-0,831]	0,975	1	<b>0,84</b>	<u>0,517</u>	<u>0,429</u>	<b>0,882</b>	<u>0,614</u>
Lésion élargie 3D*	0,7 [0,58-0,821]	0,826	2	<b>0,64</b>	<u>0,707</u>	<u>0,485</u>	<b>0,82</b>	0,687
Lésion évolutive	0,681 [0,551-0,811]	0,648	9	<b>0,48</b>	<u>0,81</u>	0,522	0,783	0,711
Lésion évolutive 3D*	0,715 [0,592-0,837]	0,978	5	<b>0,64</b>	<u>0,707</u>	<u>0,485</u>	<b>0,82</b>	0,687
Volume nouvelles lésions	0,67 [0,537-0,802]	0,52	198	<b>0,52</b>	<u>0,776</u>	0,5	0,789	0,699
Volume nouvelles lésions 3D*	0,687 [0,559-0,816]	0,68	67	<b>0,72</b>	<u>0,603</u>	<u>0,436</u>	<b>0,818</b>	<u>0,639</u>
Volume lésions élargies	0,759 [0,644-0,875]	0,566	181	<b>0,72</b>	<u>0,759</u>	0,563	<b>0,863</b>	0,747
Volume lésions élargies 3D*	0,746 [0,626-0,866]	0,689	181	<b>0,72</b>	<u>0,759</u>	0,563	<b>0,863</b>	0,747
Volume évolutif	0,758 [0,641-0,875]	0,545	303	<b>0,76</b>	<u>0,69</u>	0,514	<b>0,87</b>	0,711
Volume évolutif 3D*	0,763 [0,649-0,877]	0,493	215	<b>0,8</b>	<u>0,655</u>	<u>0,488</u>	<b>0,881</b>	0,687
Ratio volume/nouvelle lésion	0,646 [0,516-0,777]	0,273	30	<b>0,56</b>	<u>0,724</u>	<u>0,467</u>	<b>0,792</b>	0,675
Ratio volume/nouvelle lésion 3D*	0,643 [0,515-0,771]	0,251	35	<b>0,56</b>	<u>0,724</u>	<u>0,467</u>	<b>0,792</b>	0,675
Ratio volume/lésion élargie	0,763 [0,648-0,879]	0,533	86	<b>0,68</b>	<u>0,741</u>	0,531	<b>0,843</b>	0,723
Ratio volume/lésion élargie 3D*	0,75 [0,63-0,871]	0,652	86	<b>0,68</b>	<u>0,741</u>	0,531	<b>0,843</b>	0,723
Ratio volume évolutif/lésion évolutive	0,771 [0,657-0,885]	0,393	74	<b>0,68</b>	<u>0,776</u>	0,567	<b>0,849</b>	0,747
Ratio volume évolutif/lésion évolutive 3D*	0,764 [0,649-0,879]	0,462	37	<b>0,88</b>	<u>0,552</u>	<u>0,449</u>	<b>0,912</b>	<u>0,639</u>
Nouvelle lésion pondérée	0,655 [0,522-0,788]	0,444	6	0,36	0,948	0,75	0,775	0,771
Lésion élargie pondérée	0,735 [0,619-0,851]	0,807	1	<b>0,76</b>	<u>0,69</u>	0,516	<b>0,827</b>	0,711

Annexe 8. Tableau des performances de différents marqueurs radiologiques PIXYL comparées à celles du nombre de lésion comptées en FUSION pour identifier la survenue d'un échec thérapeutique entre les deux IRM (\*3D = taille seuil  $\geq$  3 mm dans les trois axes ; Volumes en mm<sup>3</sup> ; Valeurs  $\geq$  à celles de la référence selon le nombre de lésions FUSION (seuil de 3) en gras et soulignées si statistiquement différentes (test de Fisher)

Marqueur radiologique	AUC [IC95%]	p	Seuil	Se	Spé	VPP	VPN	Exact
Nouvelle lésion FUSION	0,624 [0,469-0,779]	1	3	0,25	0,881	0,333	0,831	0,759
Nouvelle lésion FUSION	0,624 [0,469-0,779]	1	2	<b>0,5</b>	0,851	<b>0,444</b>	0,877	<b>0,783</b>
Nouvelle lésion	0,592 [0,439-0,745]	0,704	3	<u>0,75</u>	<u>0,507</u>	0,267	0,895	<u>0,554</u>
Nouvelle lésion 3D*	0,574 [0,413-0,734]	0,516	3	<b>0,625</b>	<u>0,582</u>	0,263	0,867	<u>0,59</u>
Lésion élargie	0,55 [0,392-0,709]	0,42	3	<b>0,438</b>	<u>0,701</u>	0,259	0,839	0,651
Lésion élargie 3D*	0,554 [0,397-0,711]	0,441	3	<b>0,438</b>	<u>0,701</u>	0,259	0,839	0,651
Lésion évolutive	0,584 [0,432-0,736]	0,636	3	<b>0,813</b>	<u>0,403</u>	0,245	0,9	<u>0,482</u>
Lésion évolutive 3D*	0,563 [0,403-0,723]	0,448	3	<b>0,688</b>	<u>0,463</u>	0,234	0,861	<u>0,506</u>
Volume nouvelles lésions	0,581 [0,42-0,742]	0,574	78	<u>0,75</u>	<u>0,507</u>	0,267	0,895	<u>0,554</u>
Volume nouvelles lésions 3D*	0,567 [0,401-0,733]	0,432	71	<b>0,688</b>	<u>0,582</u>	0,282	0,886	<u>0,602</u>
Volume lésions élargies	0,578 [0,412-0,744]	0,606	417	<b>0,563</b>	0,746	<b>0,346</b>	0,877	0,711
Volume lésions élargies 3D*	0,582 [0,418-0,746]	0,634	417	<b>0,563</b>	0,746	<b>0,346</b>	0,877	0,711
Volume évolutif	0,615 [0,456-0,774]	0,904	322	<b>0,625</b>	<u>0,627</u>	0,286	0,875	0,627
Volume évolutif 3D*	0,604 [0,44-0,768]	0,799	491	<b>0,563</b>	<u>0,701</u>	0,31	0,87	0,675
Ratio volume/nouvelle lésion	0,544 [0,373-0,715]	0,221	31	<b>0,5</b>	<u>0,701</u>	0,267	0,849	0,639
Ratio volume/nouvelle lésion 3D*	0,564 [0,4-0,729]	0,373	39	<b>0,5</b>	<u>0,701</u>	0,269	0,842	0,663
Ratio volume/lésion élargie	0,581 [0,413-0,75]	0,636	311	<b>0,375</b>	0,866	<b>0,4</b>	0,853	<b>0,771</b>
Ratio volume/lésion élargie 3D*	0,585 [0,418-0,752]	0,664	311	<b>0,375</b>	0,866	<b>0,4</b>	0,853	<b>0,771</b>
Ratio volume évolutif/lésion évolutive	0,6 [0,443-0,757]	0,748	66	<b>0,625</b>	<u>0,657</u>	0,294	0,878	0,639
Ratio volume évolutif/lésion évolutive 3D*	0,61 [0,456-0,763]	0,85	78	<b>0,625</b>	<u>0,657</u>	0,29	0,865	0,651
Nouvelle lésion pondérée	0,599 [0,45-0,747]	0,766	2	<u>0,75</u>	<u>0,597</u>	0,308	0,909	0,627
Lésion élargie pondérée	0,549 [0,389-0,708]	0,389	0,9	<b>0,563</b>	<u>0,642</u>	0,258	0,846	0,627

Annexe 9. Tableau des performances de différents marqueurs radiologiques PIXYL comparées à celles du nombre de lésion comptées en FUSION pour identifier la survenue d'un échec thérapeutique après l'IRM de FUSION (\*taille seuil  $\geq 3$  mm dans les trois axes ; Volumes en mm<sup>3</sup> ; Valeurs  $\geq$  à celles de la référence selon le nombre de lésions FUSION (seuil de 3) en gras et soulignées si statistiquement différentes (test de Fisher)

Variable	Seuil de la variable	AUC [IC95%]	Nombre seuil de lésions optimal	Se	Spé
Certitude nouvelle lésion	0,75	0,667 [0,543-0,792]	1	<b>0,68</b>	<u>0,6</u>
Certitude lésion élargie	0,7	<b>0,746 [0,634-0,858]</b>	1	<b>0,76</b>	<u>0,71</u>
Taille nouvelle lésion en xyz	3	0,689 [0,561-0,817]	5	<b>0,48</b>	<u>0,83</u>
Taille lésion élargie en xyz	5	0,716 [0,595-0,837]	2	<b>0,56</b>	0,88
Taille nouvelle lésion en x	4	0,677 [0,539-0,815]	2	<b>0,64</b>	<u>0,71</u>
Taille lésion élargie en x	6	<b>0,731 [0,614-0,847]</b>	2	<b>0,52</b>	0,9
<b>Taille nouvelle lésion en y</b>	6	<b>0,755 [0,642-0,868]</b>	1	<b>0,76</b>	<u>0,67</u>
Taille lésion élargie en y	5	<b>0,73 [0,613-0,847]</b>	2	<b>0,64</b>	<u>0,78</u>
Taille nouvelle lésion en z	5	0,65 [0,52-0,78]	2	<b>0,44</b>	<u>0,81</u>
Taille lésion élargie en z	5	<b>0,728 [0,608-0,848]</b>	2	<b>0,6</b>	0,86
Volume nouvelle lésion	20	0,687 [0,555-0,818]	7	0,36	0,95
Volume lésion élargie	100	<b>0,734 [0,621-0,848]</b>	1	<b>0,72</b>	<u>0,72</u>

Annexe 10. Seuils lésionnels PIXYL optimaux considérés par rapport au compte lésionnel pour identifier un échec entre les deux IRM (Tailles en mm et Volumes en mm<sup>3</sup> ; Se : sensibilité ; Spé : spécificité)

Variable	Seuil de la variable	AUC [IC95%]	Volume seuil optimal	Se	Spé
Taille nouvelle lésion en x	3	0,68 [0,55-0,81]	108	<b>0,64</b>	<u>0,67</u>
Taille lésion élargie en x	0	<b>0,759 [0,644-0,875]</b>	181	<b>0,72</b>	<u>0,76</u>
Taille nouvelle lésion en y	6	0,716 [0,601-0,832]	11	<b>0,76</b>	<u>0,67</u>
Taille lésion élargie en y	0	<b>0,759 [0,644-0,875]</b>	181	<b>0,72</b>	<u>0,76</u>
Taille nouvelle lésion en z	3	0,672 [0,539-0,805]	193	<b>0,52</b>	<u>0,78</u>
Taille lésion élargie en z	0	<b>0,759 [0,644-0,875]</b>	181	<b>0,72</b>	<u>0,76</u>
Volume nouvelle lésion	20	0,674 [0,542-0,805]	60	<b>0,68</b>	<u>0,66</u>
Volume lésion élargie	30	<b>0,761 [0,647-0,875]</b>	538	<b>0,6</b>	0,86

Annexe 11. Seuils lésionnels PIXYL optimaux considérés par rapport au volume lésionnel pour identifier un échec entre les deux IRM (Tailles en mm et Volumes en mm<sup>3</sup> ; Se : sensibilité ; Spé : spécificité)

Variable	Seuil de la variable	AUC [IC95%]	Nombre seuil de lésions optimal	Se	Spé
Certitude nouvelle lésion	0,8	<b>0,663 [0,532-0,794]</b>	1	<b>0,69</b>	<u>0,7</u>
Certitude lésion élargie	0,85	<b>0,635 [0,505-0,766]</b>	1	<b>0,44</b>	0,85
Taille nouvelle lésion en xyz	0	0,592 [0,439-0,745]	3	<b>0,75</b>	<u>0,51</u>
Taille lésion élargie en xyz	11	<b>0,631 [0,499-0,764]</b>	1	<b>0,44</b>	0,84
Taille nouvelle lésion en x	5	0,594 [0,445-0,743]	1	<b>0,63</b>	<u>0,55</u>
Taille lésion élargie en x	11	<b>0,643 [0,509-0,777]</b>	1	<b>0,5</b>	0,81
<b>Taille nouvelle lésion en y</b>	6	<b>0,625 [0,485-0,764]</b>	1	<b>0,69</b>	<u>0,6</u>
Taille lésion élargie en y	15	0,596 [0,459-0,732]	1	<b>0,44</b>	0,76
Taille nouvelle lésion en z	5	0,623 [0,473-0,773]	2	<b>0,44</b>	0,78
Taille lésion élargie en z	9	<b>0,633 [0,494-0,773]</b>	1	<b>0,56</b>	<u>0,72</u>
Volume nouvelle lésion	20	0,604 [0,446-0,761]	2	<b>0,63</b>	<u>0,64</u>
Volume lésion élargie	360	<b>0,638 [0,505-0,771]</b>	1	<b>0,5</b>	0,81

Annexe 12. Seuils lésionnels PIXYL optimaux considérés par rapport au compte lésionnel pour identifier un échec postérieur à l'IRM de FUSION (Tailles en mm et Volumes en mm<sup>3</sup> ; Se : sensibilité ; Spé : spécificité)

Variable	Seuil de la variable	AUC [IC95%]	Volume seuil optimal	Se	Spé
Taille nouvelle lésion en x	5	0,585 [0,438-0,732]	73	<b>0,56</b>	<u>0,67</u>
Taille lésion élargie en x	11	<b>0,639 [0,506-0,772]</b>	360	<b>0,5</b>	0,84
Taille nouvelle lésion en y	6	<b>0,624 [0,484-0,764]</b>	33	<b>0,69</b>	<u>0,66</u>
Taille lésion élargie en y	20	0,597 [0,469-0,725]	390	<b>0,38</b>	0,88
Taille nouvelle lésion en z	5	0,604 [0,457-0,75]	66	<b>0,56</b>	<u>0,7</u>
Taille lésion élargie en z	9	<b>0,632 [0,493-0,772]</b>	333	<b>0,56</b>	0,76
Volume nouvelle lésion	30	0,59 [0,444-0,737]	34	<b>0,69</b>	<u>0,57</u>
Volume lésion élargie	240	<b>0,634 [0,497-0,771]</b>	124	<b>0,56</b>	<u>0,73</u>

Annexe 13. Seuils lésionnels PIXYL optimaux considérés par rapport au volume lésionnel pour identifier un échec postérieur à l'IRM de FUSION (Tailles en mm et Volumes en mm<sup>3</sup> ; Se : sensibilité ; Spé : spécificité)

## SUMMARY

**Introduction.** The MRI evolution of white matter lesions is a good marker for the control of multiple sclerosis (MS) by treatments with a significant threshold of three. However, the interpretation of MRI requires time-consuming expertise and becomes more complex with the importance of the lesion burden. Semi-automatic methods such as image FUSION can help the radiologist detect active lesions. Automated segmentation methods such as PIXYL are being developed to characterize lesions objectively and quickly. The main objective of this study was to compare the performance of FUSION and PIXYL methods to identify treatment failures.

**Methods.** This was a retrospective single-center study on brain MRI of MS patients followed at the Rouen University Hospital. Treatment failure was defined as the occurrence of clinical activity, relapse or disability worsening, or a change in treatment for inefficacy, from the initial MRI up to one year after follow-up MRI. The area under the sensitivity/specificity curve to identify patients with treatment failure (AUC) was the main statistical indicator used to compare the two methods.

**Results.** Treatment failure was observed in 33/83 patients (40%). The distribution of the radiological markers FUSION and PIXYL differed between the groups with and without failure. The PIXYL markers of counts and volumes of new or enlarged lesions were studied. The AUC for the number of new lesions by the two methods did not differ. The optimal sensitivity/specificity thresholds for the number of new lesions were two for FUSION and five for PIXYL rather than the usual benchmark of three new lesions. The combination of markers FUSION and PIXYL in a logistic regression model improved the AUC by 25% (0.836) and doubled the sensitivity (64%), making it possible to identify 33% of additional patients with treatment failure for a conserved specificity (96%). The number of new infratentorial lesions was the most contributing marker of PIXYL. PIXYL segmentation masks reading suggested to remain cautious due to the presence of fake active lesions in 76% of patients.

**Conclusion.** This study was the first clinical evaluation of an automated segmentation method. Classification performances of PIXYL markers synthesized by AUC sensitivity/specificity were not superior to FUSION in identifying patients with treatment failure in the year after follow-up MRI. Any significant gain in sensitivity came at the cost of a loss of specificity incompatible with clinical use. However, the combination of FUSION with PIXYL resulted in a logistic regression model whose performance improved that of FUSION alone. However, caution applies to the conclusions of this exploratory study. In fact, PIXYL segmentation masks reading found many false positives, including contrast artefacts of doubtful significance.

**Key-words:** multiple sclerosis, follow-up, MRI, automated segmentation, FUSION, PIXYL

## RÉSUMÉ

**Introduction.** L'évolutivité IRM des lésions de la substance blanche est un bon marqueur du contrôle de la sclérose en plaques (SEP) par les traitements avec un seuil significatif de trois. Cependant, l'interprétation des IRM requiert une expertise chronophage et se complexifie avec l'importance de la charge lésionnelle. Des méthodes semi-automatiques comme la FUSION des images peuvent aider le radiologue à détecter les lésions évolutives. Des méthodes de segmentation automatique comme PIXYL sont en cours de développement pour caractériser objectivement et rapidement les lésions. L'objectif de cette étude était de comparer les performances des méthodes de FUSION et PIXYL pour identifier les échecs thérapeutiques.

**Méthodes.** Il s'agissait d'une étude monocentrique rétrospective sur les IRM cérébrales de patients SEP suivis au CHU de Rouen. L'échec thérapeutique était défini par la survenue d'une activité clinique, poussée ou aggravation du handicap, ou d'un changement de traitement pour inefficacité, à partir de l'IRM initiale jusqu'à un an de l'IRM de suivi. L'aire sous la courbe sensibilité / spécificité pour identifier les patients avec échec thérapeutique (AUC) était l'indicateur statistique principal utilisé pour comparer les deux méthodes.

**Résultats.** Un échec thérapeutique était observé chez 33 / 83 patients (40 %). La répartition des marqueurs radiologiques FUSION et PIXYL différait entre les groupes avec et sans échec. Les marqueurs PIXYL de comptes et volumes des lésions nouvelles ou élargies ont été étudiés. Les AUC du nombre de nouvelles lésions selon les deux méthodes ne différaient pas. Les seuils optimaux sensibilité / spécificité du nombre de nouvelles lésions étaient de deux pour FUSION et cinq pour PIXYL plutôt que la référence habituelle de trois nouvelles lésions. L'association de marqueurs FUSION et PIXYL dans un modèle de régression logistique améliorait l'AUC de 25 % (0,836) et doublait la sensibilité (64 %), permettant d'identifier 33 % de patients supplémentaires avec échec thérapeutique pour une spécificité conservée (96 %). Le nombre de nouvelles lésions infratentorielles était le marqueur PIXYL le plus contributif. La relecture des masques de segmentation PIXYL invitait à rester prudent en raison de la présence de fausses lésions évolutives chez 76 % des patients.

**Conclusion.** Cette étude était la première évaluation clinique d'une méthode de segmentation automatique. Les performances de classification des marqueurs PIXYL synthétisées par l'AUC sensibilité / spécificité n'étaient pas supérieures à FUSION pour identifier les patients avec échec thérapeutique dans l'année suivant l'IRM de FUSION. Tout gain de sensibilité notable se faisait au prix d'une perte de spécificité incompatible avec un usage clinique. Néanmoins, la combinaison de FUSION avec PIXYL a abouti à un modèle de régression logistique dont les performances amélioraient celles de la FUSION seule. Cependant, la prudence s'applique aux conclusions de cette étude exploratoire. En effet, la relecture des masques de segmentation PIXYL retrouvait beaucoup de faux positifs dont des artefacts de contraste de signification douteuse.

**Mots clés :** sclérose en plaques, suivi, IRM, segmentation automatique, FUSION, PIXYL